# Mutations in DONSON disrupt replication fork stability and cause microcephalic dwarfism

Reynolds, John; Higgs, Martin; Zlatanou, Anastasia; Vernet, Audrey; Mottram, Rachel; Brean, Alexander ; Taylor, Malcolm; Alkuraya, Fowzan S ; Mathew, Christopher G ; Jackson, Andrew P; Stewart, Grant

*DOI:*
10.1038/ng.3790

# Mutations in *DONSON* disrupt replication fork stability and cause microcephalic dwarfism.

John J Reynolds[1,36], Louise S Bicknell[2,3,36], Paula Carroll[2,36], Martin R Higgs[1,36], Ranad Shaheen[4,36], Jennie E Murray[2], Dimitrios K Papadopoulos[5], Andrea Leitch[2], Olga Murina[2], Žygimantė Tarnauskaitė[2], Sarah R Wessel[6], Anastasia Zlatanou[1], Audrey Vernet[1], Alex von Kriegsheim[2], Rachel MA Mottram[1], Clare V Logan[2], Hannah Bye[7], Yun Li[8], Alexander Brean[1], Sateesh Maddirevula[4], Rachel C Challis[2], Kassiani Skouloudaki[5], Agaadir Almoisheer[4], Hessa S Alsaif[4], Ariella Amar[7], Natalie J Prescott[7], Michael B Bober[9], Angela Duker[9], Eissa Faqeih[10], Mohammed Zain Seidahmed[11], Saeed Al Tala[12], Abdulrahman Alswaid[13], Saleem Ahmed[14,15], Jumana Yousuf Al-Aama[14,15], Janine Altmüller[16], Mohammed Al Balwi[17], Angela F Brady[18], Luciana Chessa[19], Helen Cox[20], Rita Fischetto[21], Raoul Heller[22], Bertram D Henderson[23], Emma Hobson[24], Peter Nürnberg[16], E Ferda Percin[25], Angela Peron[26,27], Luigina Spaccini[26], Alan J Quigley[28], Seema Thakur[29], Carol A Wise[30], Grace Yoon[31,32], Maha Alnemer[33], Pavel Tomancak[5], Gökhan Yigit[8], A Malcolm R Taylor[1], Martin AM Reijns[2], Michael A Simpson[7], David Cortez[6], Fowzan S Alkuraya[4,35,*], Christopher G Mathew[7,34,35,*], Andrew P Jackson[2,35,*] & Grant S Stewart[1,35,*].

[1]Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK.

[2]MRC Human Genetics Unit, IGMM, University of Edinburgh, Edinburgh, UK.

[3]Present address: Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand.

[4]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia.

[5]Max-Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany.

[6]Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.

[7]Department of Medical and Molecular Genetics, Faculty of Life Science and Medicine, King's College London, 7th Floor Tower Wing, Guy's Hospital, London, UK.

[8]Institute of Human Genetics, University Medical Center Göttingen, Göttingen, Germany.

[9]Nemours/Alfred I. duPont Hospital for Children, Wilmington, Delaware, USA.

[10]Department of Pediatric Subspecialties, Children's Hospital, King Fahad Medical City, Riyadh, Saudi Arabia.

[11]Department of Pediatrics, Security Forces Hospital, Riyadh, Saudi Arabia.

[12]Armed forces hospital, SR. P.D. Genetic unit. Khamis Mushayt, Saudi Arabia.

[13]Department of Pediatrics, King Abdulaziz Medical City, Riyadh, Saudi Arabia.

[14]Department of Genetic Medicine, King Abdulaziz University Hospital, Jeddah, Saudi Arabia

[15]Princess Al-Jawhara Al-Brahim Center of Excellence in Research of Hereditary Disorders, Jeddah, Saudi Arabia.

[16]Cologne Center for Genomics and Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany.

[17]Department of Pathology and Laboratory Medicine, King Abdulaziz Medical City, National Guard Health Affairs, Riyadh, Saudi Arabia.

[18]North West Thames Regional Genetics Service, London North West Healthcare NHS Trust, Harrow, UK.

[19]Department of Clinical and Molecular Medicine, University La Sapienza, Roma, Italy.

[20]West Midlands Regional Clinical Genetics Service, Birmingham Women's Hospital, West Midlands, UK.

[21]Pediatric Hospital Giovanni XXIII, Bari, Italy.

[22]Institute of Human Genetics, University Hospital of Cologne, Cologne, Germany.

[23]Divison of Clinical Genetics, Faculty of Health Sciences, UFS, Bloemfontein, South Africa.

[24]Department of Genetics, Yorkshire Regional Genetic service, Chapel Allerton Hospital, Leeds, UK.

[25]Department of Medical Genetics, Gazi University Faculty of Medicine, Besevler Ankara, Turkey.

[26]Clinical Genetics Unit, Division of Maternal Fetal Medicine, Department of Obstetrics and Gynecology, V. Buzzi Children's Hospital, Universita' degli Studi di Milano, Milan, Italy.

[27]Child Neuropsychiatry Unit – Epilepsy Center, San Paolo University Hospital, Department of Health Sciences, Università degli Studi di Milano, Milan, Italy.

[28]Dept of Radiology, Royal Hospital for Sick Children, Edinburgh, UK.

[29]Dept of Genetic and Fetal Medicine, Fortis lafemme, New Delhi, India.

[30]Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, Texas, USA.

[31]Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, University of Toronto, Toronto, Canada.

[32]Department of Pediatrics, Division of Neurology, The Hospital for Sick Children, University of Toronto, Toronto, Canada.

[33]Department of Obstetrics and Gynecology, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia.

[34]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, South Africa.

[35]These authors jointly directed this work.

[36]These authors contributed equally to this work.

79 *Correspondence to:

80 GSS, g.s.stewart@bham.ac.uk; APJ, Andrew.jackson@igmm.ed.ac.uk; CGM,

81 christopher.mathew@kcl.ac.uk; FSA, falkuraya@kfshrc.edu.sa

**Abstract**

To ensure efficient genome duplication, cells have evolved numerous factors that promote unperturbed DNA replication, and protect, repair and restart damaged forks.

Here we identify DONSON as a novel fork protection factor, and report biallelic *DONSON* mutations in 29 individuals with microcephalic dwarfism. We demonstrate that DONSON is a replisome component that stabilises forks during genome replication. Loss of DONSON leads to severe replication-associated DNA damage arising from nucleolytic cleavage of stalled replication forks. Furthermore, ATR-dependent signalling in response to replication stress is impaired in DONSON-deficient cells, resulting in decreased checkpoint activity, and potentiating chromosomal instability. Hypomorphic mutations substantially reduce DONSON protein levels and impair fork stability in patient cells, consistent with defective DNA replication underlying the disease phenotype.

In summary, we identify mutations in *DONSON* as a common cause of microcephalic dwarfism, and establish DONSON as a critical replication fork protein required for mammalian DNA replication and genome stability.

Microcephalic primordial dwarfism (MPD) is the collective term for a group of human disorders characterised by intra-uterine and postnatal growth delay alongside marked microcephaly[1], and includes disorders such as MOPD II, ATR/ATRIP-Seckel syndrome and Meier-Gorlin syndrome. Mutations in genes encoding either components of the DNA replication machinery (replisome) or genome stability proteins are a frequent cause of microcephalic dwarfism[2-14].

During the course of normal DNA replication, a subset of replication forks may stall, causing 'replication stress'[15]. This stalling can be caused by endogenous or exogenous sources, such as collision of the replisome with DNA lesions or the transcriptional machinery, or replication of difficult to replicate genomic regions. To facilitate efficient genome duplication, stalled replication forks must be stabilised and protected from collapse. Multiple factors safeguard replication fork stability, many of which function within the ATR-CHK1-dependent replication stress response[16-18]. This pathway ensures that fork stabilisation is tightly coordinated with a global reduction in DNA synthesis, allowing stalled or damaged forks to be repaired and restarted[19,20].

Exome sequencing analysis of microcephalic dwarfism patients has identified several novel factors that regulate replication and/or the replication stress response. Using this strategy, we recently identified mutations in *TRAIP* in individuals with MPD[5], and demonstrated that TRAIP is required for the response to replication-blocking DNA lesions. To identify similar disease-associated genes, we carried out whole exome sequencing of genetically uncharacterised patients with microcephaly. Here, we report the identification of *DONSON* as a new microcephalic dwarfism gene, and demonstrate that DONSON is a novel replisome component that maintains genome stability by protecting stalled/damaged replication forks.

## Results

### *DONSON* mutations identified in microcephalic dwarfism patients

Whole exome sequencing (WES) was undertaken on 26 patients with microcephaly and reduced stature. After aligning WES reads to the reference genome, variant calling, and filtering for rare variants (MAF <0.005), analysis under a recessive model of inheritance identified rare biallelic variants in the *DONSON* (‘*Downstream neighbour of SON*’) gene in nine affected individuals from seven families. Sanger capillary sequencing confirmed the presence of these mutations in these patients (P1-1 to P3, P6, and P9 to P11, **Table 1**). Subsequent re-sequencing of an additional 230 patients with primary microcephaly, microcephaly with reduced stature, or MPD, identified five additional families with compound heterozygous mutations in *DONSON* (P4, P5, P7, P8, P12; **Table 1**). All variants segregated amongst family members in a manner consistent with an autosomal recessive trait, and were present at a frequency of <0.5% in the ExAC database[21].

Two other concurrent molecular genetic studies provided further independent evidence to support the identification of *DONSON* as a novel human disease gene. Firstly, exome sequencing was carried out on a consanguineous Palestinian family previously reported to have a Fanconi Anaemia-like disorder[22]. These patients presented with microcephaly, short stature, slow growth and forearm and thumb dysplasia, although no individuals had haematological evidence of bone marrow failure. This WES analysis revealed a deleterious homozygous transition, c.1337T>C, resulting in substitution of a highly conserved residue (p.M446T) in all three affected individuals (P13-1, P13-2, P13-3; **Table 1**, **Supplementary Fig. 1**). Secondly, a study of five consanguineous families in Saudi Arabia with extreme microcephaly and short stature allowed a 1.6 Mb haplotype shared by all five families (combined multipoint LOD score $Z$= 8.0) to be mapped to a defined critical interval on chromosome 21 that contained *DONSON*. Whole exome and

genome sequencing identified a single rare variant at this locus in *DONSON,* c.786-22A>G. Capillary sequencing confirmed this intronic variant to be homozygous in all seven affected individuals from this study (P14 to P18-3; **Table 1**), identical to that detected in two Saudi Arabian individuals present within the first study described above (P11, P12).

Subsequently, a further five individuals from three different families with *DONSON* mutations were identified in additional MPD patients recruited to two of the genetic studies described above (P19 to P21-2; **Table 1**).

### ***DONSON* mutations give rise to severe microcephaly with short stature**

Despite their identification in separate studies, all patients with *DONSON* mutations had similar clinical phenotypes. Marked microcephaly was present (OFC -7.5 +/- 2.4 SD), with a substantial reduction in cerebral cortical size, along with decreased gyral folding evident on neuroimaging (**Fig. 1a** and **Supplementary Fig. 2**), similar to that previously reported for other primary microcephaly and microcephalic dwarfism patients[5,23-25]. Height was reduced (-3.2 +/- 1.4 SD), although much less so than head circumference (**Fig. 1a**), and to a lesser degree than observed in other microcephalic dwarfism-associated disorders (where height was typically ≤ -4 SD)[2,3,5,8-10,24,26]. Minor skeletal abnormalities were present in several patients, including fifth finger clinodactyly, syndactyly, brachydactyly, hypoplasia of carpal/metacarpal/phalangeal bones, or radial head dislocation (**Supplementary Table 1**). Absent/hypoplastic patellae were present in patients P12, P20-1 and P20-2. Notably, patient P19 had bilateral hypoplasia of the radius and thumb, which, together with the limb abnormalities displayed by P13-1 and 13-2, established radial ray defects as an uncommon but recurrent phenotype. In family P21, the most extreme phenotype was observed, with substantial limb shortening/reduction in association with foetal lethality (**Supplementary Fig. 3**). Aside from microcephaly, neither a recognisable facial phenotype (**Fig. 1b**) nor recurrent malformations affecting other organ systems were evident. Intellectual disability, if present, was typically mild.

In conclusion, the number of biallelic variants identified, combined with a common clinical presentation, provided strong evidence for *DONSON* being a novel human disease gene, associated with autosomal recessive inheritance. We therefore investigated the consequence of these mutations on DONSON protein function.

### *DONSON* mutations markedly reduce protein levels

*DONSON* mutations were identified in 29 individuals, and comprised a range of mutation classes (**Fig. 1c**). Notably, no biallelic nonsense or frameshift mutations were observed, indicating that mutations likely resulted in partial loss of DONSON function. To investigate this, we established patient-derived primary fibroblast or lymphoblastoid cell lines representing a range of mutations. Immunoblotting demonstrated marked decreases in DONSON protein levels in all cell lines tested (**Fig. 2a, b**), establishing that the analysed mutations affected DONSON protein expression.

Mutations in multiple families were associated with two different ancestral haplotypes (P1-1 to P7; P11, P12 and P14 to 18-3 respectively; **Table 1**) and were investigated in more detail. Firstly, as described above, nine individuals (P11, P12 and P14 to P18-3) were homozygous for the c.786-22A>G mutation, predicted to enhance a cryptic splice donor site within intron 4 (MaxEntScan)[27]. Consistent with a common ancestral founder, five consanguineous families of Saudi Arabian origin bearing this mutation (P14 to P18-3) shared a 1.6 Mb haplotype region of chromosome 21 (**Supplementary Fig. 4**). qRT-PCR analysis of RNA isolated from four patients with this variant demonstrated a significant decrease in full-length transcript, and increased skipping of exon 5 (**Supplementary Fig. 5**). This resulted in an out-of-frame mRNA predicted to undergo NMD, explaining the substantial reduction in protein levels seen in fibroblasts homozygous for this mutation (P12; **Fig 2b**).

Secondly, two missense variants in DONSON, p.S28R and p.K489T, and an intronic variant (c.786-33A>G) were present in seven individuals of European ancestry and one Somali (P1-1 to P7; **Table 1**). These were associated with a different ancestral haplotype, comprising a shared 127.7 kb genomic region (**Supplementary Table 2**). No other deleterious biallelic variants were present in the four other genes within this region. Despite the close proximity of the c.786-33A>G intronic variant to the Saudi Arabian c.786-22A>G mutation, the former change did not disrupt splicing between exons 4 and 5, as assessed by either mini gene splicing assays (**Supplementary Fig. 6**), or RT-PCR analyses of patient cell lines (data not shown), indicating this variant is unlikely to be pathogenic. Since DONSON protein levels were severely reduced in cells from patients inheriting this haplotype allele in combination with a truncating mutation (P2, P6), this suggested that either one or both of the missense variants associated with this haplotype (p.S28R, p.K489T) compromised protein levels (**Fig. 2b**).

To investigate these two variants, we established isogenic HeLa cell lines expressing doxycycline-inducible, siRNA-resistant wild type (WT) or mutant (p.S28R, p.K489T) GFP-tagged DONSON. Following siRNA depletion of endogenous DONSON (**Supplementary Fig. 7**), and induction of exogenous GFP-DONSON, immunoblotting revealed that the p.K489T mutation, but not p.S28R, reduced protein levels in a post-transcriptional manner (**Fig. 2c** and **Supplementary Fig. 8**). This suggested that the p.K489T substitution within the second haplotype causes the decreased DONSON protein levels observed in P2 and P6 (**Fig. 2b**). The K489T variant is present as a rare allele in the population, observed at a frequency of 0.00099 in the ExAC database[21]. In the patients reported here, it is always observed *in trans* with a frameshift or other protein-disrupting allele, suggesting that it is a functionally weak variant insufficient to cause disease alone, a conclusion supported by the presence of a single homozygous individual in ExAC.

Eight other point mutations were identified in patients: six missense substitutions, a two amino-acid deletion and an amino-acid insertion (**Table 1** and **Fig. 1c**). Notably these were at highly conserved residues (**Supplementary Fig. 1**) and predicted to be deleterious (Alamut Visual). In agreement, exogenous expression of the p.M446T mutant resulted in significantly reduced protein levels (**Fig. 2c**), similar to cells derived from patients P13-1, P13-2 and P13-3. Furthermore, five of these mutations disrupted the subcellular localisation of GFP-DONSON (**Supplementary Fig. 9**), suggesting that these alterations also compromise DONSON protein function via protein mis-localisation.

Finally, an intronic mutation, c.1047-9A>G, was present in three individuals (P9, P21-1 and P21-2). qRT-PCR analysis of RNA isolated from the two patients homozygous for this variant (P21-1, P21-2) revealed a substantial reduction in *DONSON* transcript levels compared to normal controls (**Supplementary Fig. 10**). This variant was also observed *in trans* with the missense mutation F292L in patient P9. Since cells derived from this individual exhibited severely reduced levels of DONSON protein (**Fig. 2b**), it is likely that this intronic change also perturbs DONSON protein expression.

Taken together, the deleterious consequences of the identified mutations on splicing, transcript abundance, subcellular localisation and/or protein levels strengthened our conclusion that *DONSON* was a novel human disease gene. The fact that knockout of murine *Donson* leads to developmental lethality[28], together with the presence of residual DONSON protein in patient-derived cell lines (**Fig. 2a**, **b** and **Supplementary Fig. 11**), supports the notion that the identified mutations are hypomorphic, retaining some residual function.

**DONSON stabilises replication forks during normal DNA replication**

While *DONSON* is highly conserved in metazoa and plants, its precise function(s) remained to be defined. *Humpty-dumpty* (*hd*), the *Drosophila* ortholog of *DONSON*, has

been proposed to play a role in cell proliferation: Hd expression peaks during S-phase; *hd* mutants have an 'egg shell' phenotype; and clonal inactivation of *hd* impairs genome replication in larval tissues[29].

In light of this, we investigated whether human DONSON might play a similar role. After synchronising cells with a double thymidine block, we observed that human DONSON, like Hd, was also maximally expressed during S-phase, mirroring Cyclin A expression (**Fig. 3a**). Furthermore, depletion of DONSON resulted in a significant increase in BrdU-positive cells observed by FACS, consistent with a role in promoting efficient S-phase progression (**Fig. 3b**). Given these data, we next used DNA fibre analysis to assess whether DONSON depletion led to decreased DNA replication fork progression. Although fork progression rates did not decrease in cells lacking DONSON (**Fig. 3c**), compromising DONSON expression increased spontaneous replication fork stalling, with a concomitant decrease in the number of ongoing forks (**Fig. 3d**). Moreover, we also observed increased replication fork asymmetry in cells depleted of DONSON, indicating replication fork instability (**Fig. 3e**). Together, this suggests that the increase in BrdU-positive DONSON-depleted cells may reflect prolonged S-phase due to stalled replication forks, rather than a global reduction in DNA synthesis.

Since DNA replication is closely linked with genome stability[15,30,31], we reasoned that loss of DONSON would lead to a failure to complete timely replication and increased S-phase DNA damage. To test this hypothesis, we combined immunofluorescence of γ-H2AX and 53BP1 (markers of DNA damage and DNA double strand breaks respectively) with EdU labelling to identify S-phase cells. We observed that a significant proportion of DONSON-depleted cells exhibited spontaneous γ-H2AX and 53BP1 foci (**Fig. 3f**), of which the majority occurred in S-phase cells (**Supplementary Fig. 12a-d**), consistent with the identification of DONSON as a potential genome stability regulator by high-throughput siRNA screening[32]. Using pulsed-field gel electrophoresis, we also observed elevated

levels of DNA double strand breaks in the absence of DONSON (**Supplementary Fig. 12e**). Taken together, these data support a role for DONSON in maintaining replication fork stability during unperturbed DNA replication, and demonstrate that spontaneous DNA damage arises in replicating cells in the absence of DONSON.

**DONSON is a component of the replisome**

To shed further light on the role of DONSON in regulating replication fork stability, we carried out mass spectrometry screening to identify interaction partners of GFP-tagged DONSON. Amongst the interactors, we detected multiple replication proteins including subunits of the MCM helicase and the GINS complex (**Fig. 4a** and **Supplementary Table 3**). To confirm these findings, we carried out pull-down analyses coupled with immunoblotting to identify GFP-DONSON binding proteins. Consistent with our mass spectrometry data, we detected interactions between GFP-DONSON and the replisome components MCM2, MCM7, Treslin and PCNA (**Fig. 4b**), suggesting that DONSON associated with the replisome.

We next used three complementary techniques to assess whether DONSON localised to sites of DNA replication. Firstly, we carried out proximity ligation assays (PLA) of GFP-DONSON with the replication proteins PCNA and RPA. In line with DONSON being closely associated with the replication machinery, we observed robust PLA signals between GFP-DONSON and both PCNA and RPA (**Fig. 4c**, **d**). We next performed Fluorescence Cross-Correlation Spectroscopy (FCCS)[33,34] in live HeLa cells stably co-expressing RFP-PCNA and GFP-DONSON, to measure the degree of co-diffusion of these molecules. Significantly increased co-diffusion of PCNA and DONSON was observed in S-phase PCNA foci, but not in nuclei of non-replicating cells (**Fig. 4e-f** and **Supplementary Fig. 13**), indicating that these proteins interacted during DNA replication. Finally, we utilised iPOND (isolation of proteins on nascent DNA)[35] combined with mass

spectrometry to ascertain whether DONSON is present on newly replicated DNA. Crucially, this approach demonstrated that DONSON, like MCMs and RPA, was significantly enriched at replication forks compared to mature chromatin (**Fig. 4g**).

Collectively, these data strongly support the conclusion that DONSON is a novel replisome component that plays a role in promoting fork stability.

**DONSON depletion impairs cell-cycle checkpoint activation**

Since our data suggested that DONSON functions to protect replication forks during unperturbed DNA replication, we extended our findings to evaluate the role of DONSON in preventing replication fork stalling following exogenous replication stress. Exposure to the replication stress-inducing agents hydroxyurea (HU) and mitomycin C (MMC) induced significantly more fork stalling in DONSON-depleted cells than in control cells (**Fig. 5a-b**). DONSON depletion also resulted in a failure to suppress new origin firing upon exogenous replication stress (**Fig. 5c**). Since suppression of new origin firing reflects checkpoint activity, this suggests that DONSON is required for efficient activation of the intra-S phase checkpoint. To further investigate this, we measured activation of this checkpoint after inhibition of ATR (VE821; ATRi), the apical kinase which governs the replication stress response[18]. Upon HU exposure and ATR inhibition, we observed no difference in the number of new origins fired between control or DONSON-depleted cells, indicating that DONSON and ATR may function within the same pathway to activate the intra-S phase checkpoint (**Fig. 5d**).

We next examined whether the ATR-dependent replication stress response was functional in the absence of DONSON. We first monitored ATR pathway activation in DONSON-depleted cells treated with HU or MMC by immunoblotting, using phospho-specific antibodies to known ATR substrates. This analysis revealed that cells lacking DONSON failed to efficiently phosphorylate a number of ATR substrates, such as CHK1

and NBS1, in response to HU or MMC (**Fig. 5e** and **Supplementary Fig. 14a**). Moreover, ATR autophosphorylation on T1989, another marker of ATR activation[36], was reduced (**Supplementary Fig. 14b**). Loss of DONSON also significantly increased mitotic indices following exposure to HU or MMC as measured by FACS, demonstrating that DONSON-depleted cells fail to efficiently activate the G2/M checkpoint in response to replication stress (**Fig. 5f** and **Supplementary Fig. 14c**). We next determined whether the reduced ATR signalling observed was due to decreased levels of RPA-coated ssDNA, which is the stimulus for ATR activation. Surprisingly, DONSON-depleted cells exhibited elevated levels of RPA-coated ssDNA following HU treatment (**Supplementary Fig. 15**), consistent with defective activation of the ATR-dependent replication stress response.

Dysregulated DNA replication combined with impaired intra-S phase checkpoint signalling, such as in ATR-deficient cells[37-41], gives rise to extensive chromosome breakage and genome instability. Consistent with this, we observed significantly elevated levels of spontaneous micronuclei and chromatid gaps/breaks in cells lacking DONSON (**Fig. 6a** and **Supplementary Fig. 16a**), which was exacerbated by exposure to HU or MMC (**Supplementary Fig. 16a-c**). We also observed spontaneously-arising highly-fragmented or completely pulverised metaphase chromosomes in cells lacking DONSON, which increased upon exogenous replication stress (**Fig. 6b** and **Supplementary Fig. 16d**).

Together, these data confirm that upon exogenous replication stress, DONSON is required to stabilise stalled replication forks, efficiently activate the intra-S and G/2M cell-cycle checkpoints, and maintain genome stability.

**Cleavage of stalled replication forks in DONSON-deficient cells**

It has been proposed that the spontaneous DNA damage arising in ATR-deficient cells is due to processing of stalled/damaged forks by SLX4-associated structure-specific

nucleases, such as MUS81, SLX1 and XPF[42-46]. We therefore postulated that the replication abnormalities and chromosomal aberrations of DONSON-deficient cells might arise via similar mechanisms. Indeed, the spontaneous replication fork asymmetry and H2AX phosphorylation exhibited by DONSON-depleted cells were partially reduced by co-depletion of either MUS81 or XPF (**Fig. 6c**, **d**). Moreover, co-depletion of MUS81 or XPF also reduced chromosome breakage and genomic pulverisation in these cells (**Fig. 6e**-**g**). From this, we conclude that the severe genome instability apparent in the absence of DONSON is due to nucleolytic processing of damaged replication forks by structure-specific nucleases.

**Replication stress-induced genomic damage in DONSON patient cells**

To link the role of DONSON in regulating replication fork stability and the phenotype of patients with *DONSON* mutations, we characterised replication dynamics and genomic stability of patient-derived fibroblasts. All DONSON patient-derived cell lines examined (P2, P6, P9, P10-2, P12) showed higher levels of spontaneous fork asymmetry and fork stalling than cells from unaffected individuals (**Supplementary Figs. 17a**, **18**). Furthermore, patient-derived cells also exhibited elevated fork asymmetry and fork stalling following HU exposure, combined with defective intra-S phase checkpoint activation (**Supplementary Fig. 17a**, **18**). Finally, levels of S-phase DNA damage and chromosome breakage were also elevated in these cell lines (**Supplementary Fig. 17b-c**). Together, these observations provide a potential pathological explanation for the clinical phenotype.

In addition, using isogenic cell lines inducibly expressing GFP-DONSON (**Fig. 2c**), we observed that expression of the haplotype-associated S28R mutant, but not the K489T variant, complemented loss of endogenous DONSON by rescuing the spontaneous fork stalling observed upon DONSON depletion (**Supplementary Fig. 19**). This is consistent with K489T being the pathogenic variant within the haplotype region (**Fig. 2c**).

Finally, we set out to demonstrate that the patient-associated cellular phenotypes were directly due to mutation of *DONSON*. We first established three paired isogenic cell lines via transduction of patient-derived fibroblasts with retroviral expression vectors encoding WT *DONSON* or an empty vector (**Fig. 7a**). Importantly, the spontaneous DNA damage, replication fork stalling, replication fork asymmetry and intra-S phase checkpoint defect were all corrected by expression of WT *DONSON* (**Fig. 7b-d** and **Supplementary Fig. 20**), confirming that these phenotypes were directly due to DONSON deficiency. Lastly, using one of these cell lines, we also observed that inhibition of ATR and mutation of *DONSON* are epistatic with regard to the observed replication abnormalities (**Supplementary Fig. 21**).

## Discussion

Here we identify *DONSON* as a novel human disease gene, and describe 29 individuals with a range of mutations in *DONSON*, establishing such alterations as a frequent cause of microcephalic dwarfism. Since normal embryonic development requires rapid cellular proliferation[47,48] it is exquisitely sensitive to genetic perturbations that impact DNA replication[1-3,6]. A failure to complete timely genome duplication will profoundly affect the number of cells generated during embryonic development. For example, hypomorphic mutations in *ATR* result in severe microcephaly and growth retardation, both in humans and in a murine model[2,3,49,50], due to the role that ATR plays in preventing replication stress during development[49-50]. We propose that mutation of *DONSON* similarly reduces the number of cells generated during development via a failure to maintain replication fork stability in the presence of endogenous replication stress, thus explaining the decreased organism size observed. Furthermore, since brain development requires rapid proliferation of neural progenitor cells within a limited timeframe[47], it is particularly sensitive to disruptive genetic perturbations. This may explain why brain development is disproportionately affected in these individuals compared to growth.

DONSON has no predicted domain structure or paralogs, and previous characterisation has been limited to two previous studies: an siRNA screen proposing that DONSON regulates genome stability, and a study of its Drosophila ortholog *Humpty-dumpty* suggesting a role in cell proliferation[29,32]. Consistent with this, we establish that DONSON is a replisome component that ensures replication fork stability, and promotes efficient activation of both intra-S and G2/M cell-cycle checkpoints upon exogenous replication stress. Loss of DONSON leads to increased spontaneous stalling of replication forks, which are subsequently cleaved into replication-associated DNA double strand breaks by structure-specific nucleases. Defective cell-cycle checkpoint activation in DONSON-deficient cells then allows transmission of these breaks into mitosis, accounting

for the elevated chromosomal damage and genome fragmentation observed (**Supplementary Fig. 22**). Additional studies will be important to confirm this model, and to investigate whether DONSON is a constitutive component of the replisome, or whether it is recruited to a subset of replication forks. Furthermore, establishing which replisome components DONSON directly interacts with, and the functional importance of these associations, will also inform understanding of its biological function.

The mechanism by which DONSON ensures replication fork stability and promotes checkpoint activation remains to be defined. We propose that in addition to being a replisome component, DONSON is also involved in promoting the ATR-CHK1 replication stress response, since we observed that DONSON-depleted cells exhibit defective activation of cell cycle checkpoints and reduced ATR-dependent signalling in response to exogenous replication stress. This hypothesis is supported by the observation that impaired replication alone, such as that arising from a hypomorphic mutation in MCM4 (*MCM4*$^{Chaos3/Chaos3}$), does not give rise to decreased CHK1 phosphorylation or increased new origin firing upon replication stress[51]. However, it is unclear whether DONSON functions directly or indirectly to regulate the ATR-CHK1 pathway. Our demonstration that cells lacking DONSON do not exhibit a global reduction in replication, or decreased levels of RPA-coated ssDNA, indicates that loss of DONSON does not affect the cells ability to generate the primary stimulus for ATR pathway activation. Based on this, we propose that either DONSON directly activates ATR, in a manner similar to TOPBP1[52] or ETAA1[53,54], or functions indirectly to regulate other factors necessary for efficient ATR-CHK1 signalling, such as the MRE11/RAD50/NBS1 (MRN)[55] complex or TIPIN/Timeless[19,20]. Since DONSON does not contain a canonical ATR activation domain, which is found in both TOPBP1 and ETAA1, we favour the latter possibility. However, how DONSON acts to promote ATR signalling is not yet clear, and future work will be critical in establishing whether this is direct or indirect.

441 It is also evident that the cellular phenotype of cells lacking DONSON cannot be
442 explained solely by abnormal DNA replication or defective ATR-dependent signalling. In
443 particular, exposure of cells lacking ATR to exogenous replication stress results in highly
444 elevated levels of H2AX phosphorylation, a situation not observed upon DONSON loss
445 (**Fig. 5e**), despite the presence of substantial amounts of DNA damage. Therefore, whilst
446 our observations are consistent with a role for DONSON in promoting ATR-CHK1
447 signalling, DONSON may also impact on other pathways that promote H2AX
448 phosphorylation at the replication fork, for example those governed by ATM or the MRN
449 complex.
450 In summary, we have identified *DONSON* as a novel disease gene that plays a key
451 role in regulating cellular replication and cell cycle checkpoints. Further studies examining
452 how DONSON functions will provide fundamental insight into how cells maintain replication
453 fork integrity, and how these pathways prevent human disease.

**Acknowledgements**

We would like to thank the families and clinicians for their involvement and participation. We are grateful to R. S. Taylor (University of Manchester) and D.-J. Kleinjan (University of Edinburgh), J.Bartek and C.Bartek (Copenhagen) for their kind gift of reagents. We thank E. Freyer, J. Wills, J. Ding, A. Fluteau, C. Keith, D. Longman and the IGMM FACS, core sequencing and mass-spectrometry facilities for technical assistance and advice. The Walking With Giants Foundation and Potentials Foundation supported the Primordial Dwarfism Registry (M.B.B.). This work is supported by funding from Cancer Research UK (C17183/A13030) (G.S.S., M.R.H. and A.V.), the Medical Research Council (MR/M009882/1) (J.J.R.), Worldwide Cancer Research (13-1012) (A.Z.), Birmingham Children's Hospital Research Foundation (BCHRF400) (R.M.A.M.), University of Birmingham (J.J.R., R.M.A.M., A.B.), Newlife Foundation for Disabled Children (P.C., L.S.B.), Medical Research Scotland (L.S.B.) and the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London (H.B., A. Amar, N.J.P, M.A.S., C.G.M.), German Federal Ministry of Education and Research (BMBF) (1GM1404; E-RARE network EuroMicro) (G.Yigit), KSCDR funding and KACST grant 09-MED941-20 (F.S.A.), Swiss National Science Foundation (P2ZHP3_158709) (O.M.), Medical Research Council, the Lister Institute for Preventative Medicine and the European Research Council (ERC, 281847) (A.P.J.).

**Author Contributions**

J.J.R., M.R.H., P.C., O.M., A.Z., A.L., R.M.A.M., A.B. and G.S.S. designed and performed the cell biology experiments. J.E.M., L.S.B., R.S, C.V.L., F.S.A, M.A.S., C.G.M., Y.L., S.M. and G. Yigit performed NGS sequencing and analysis. L.S.B., P.C., R.C.C., R.S., A.V., J.E.M., M.A.S., C.V.L., Z.T., M.A.M.R., H.B., A. Amar., S.M., A. Almoisheer, H.S.A. and N.J.P. performed sequencing, genotyping, linkage analysis, analysis of splicing and other

molecular genetics experiments. D.C. and S.R.W. performed iPOND experiments. P.T., D.K.P. and K.S. performed FCCS analysis. A.V.K. performed mass spectrometry analysis. E.F., M.Z.S., S.A.T., A. Alswaid, S.A., J.Y.A., M.A.B., A.F.B., L.C., H.C., A.D., R.F., E.H., E.F.P., A.P., L.S., S.T., G. Yoon., J.A., P.N., A.J.Q., B.D.H., M.A. and R.H. contributed clinical cases and clinical data and analysis for the study. M.B.B., C.A.W., J.E.M., L.S.B, A.M.R.T., F.S.A., C.G.M. and A.P.J. recruited study cohorts, and performed review of phenotypes and sample collection. J.J.R., M.R.H., L.S.B., A.P.J. and G.S.S. wrote the manuscript. The study was planned and supervised by G.S.S., C.G.M., F.S.A and A.P.J.

**Competing Financial Interests Statement**

The authors declare no competing financial interests.

## References


1. Klingseisen, A. & Jackson, A.P. Mechanisms and pathways of growth failure in primordial dwarfism. *Genes Dev* **25**, 2011-24 (2011).
2. O'Driscoll, M., Ruiz-Perez, V.L., Woods, C.G., Jeggo, P.A. & Goodship, J.A. A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. *Nat Genet* **33**, 497-501 (2003).
3. Ogi, T. *et al.* Identification of the first ATRIP-deficient patient and novel mutations in ATR define a clinical spectrum for ATR-ATRIP Seckel Syndrome. *PLoS Genet* **8**, e1002945 (2012).
4. German, J. Bloom's syndrome. I. Genetical and clinical observations in the first twenty-seven patients. *Am J Hum Genet* **21**, 196-227 (1969).
5. Harley, M.E. *et al.* TRAIP promotes DNA damage response during genome replication and is mutated in primordial dwarfism. *Nat Genet* **48**, 36-43 (2016).
6. Qvist, P. *et al.* CtIP Mutations Cause Seckel and Jawad Syndromes. *PLoS Genet* **7**, e1002310 (2011).
7. Rosin, N. *et al.* Mutations in XRCC4 cause primary microcephaly, short stature and increased genomic instability. *Hum Mol Genet* **24**, 3708-17 (2015).
8. Bicknell, L.S. *et al.* Mutations in the pre-replication complex cause Meier-Gorlin syndrome. *Nat Genet* **43**, 356-9 (2011).
9. Bicknell, L.S. *et al.* Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. *Nat Genet* **43**, 350-5 (2011).
10. Guernsey, D.L. *et al.* Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome. *Nat Genet* **43**, 360-4 (2011).

11. Fenwick, A.L. *et al.* Mutations in CDC45, Encoding an Essential Component of the Pre-initiation Complex, Cause Meier-Gorlin Syndrome and Craniosynostosis. *Am J Hum Genet* **99**, 125-38 (2016).
12. Murray, J.E. *et al.* Extreme growth failure is a common presentation of ligase IV deficiency. *Hum Mutat* **35**, 76-85 (2014).
13. Murray, J.E. *et al.* Mutations in the NHEJ component XRCC4 cause primordial dwarfism. *Am J Hum Genet* **96**, 412-24 (2015).
14. Shaheen, R. *et al.* Genomic analysis of primordial dwarfism reveals novel disease genes. *Genome Res* **24**, 291-9 (2014).
15. Zeman, M.K. & Cimprich, K.A. Causes and consequences of replication stress. *Nat Cell Biol* **16**, 2-9 (2014).
16. Zou, L. & Elledge, S.J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542-8 (2003).
17. MacDougall, C.A., Byun, T.S., Van, C., Yee, M.C. & Cimprich, K.A. The structural determinants of checkpoint activation. *Genes Dev* **21**, 898-903 (2007).
18. Nam, E.A. & Cortez, D. ATR signalling: more than meeting at the fork. *Biochem J* **436**, 527-36 (2011).
19. Chou, D.M. & Elledge, S.J. Tipin and Timeless form a mutually protective complex required for genotoxic stress resistance and checkpoint function. *Proc Natl Acad Sci U S A* **103**, 18143-7 (2006).
20. Kemp, M.G. *et al.* Tipin-replication protein A interaction mediates Chk1 phosphorylation by ATR in response to genotoxic stress. *J Biol Chem* **285**, 16562-71 (2010).
21. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).

22. Milner, R.D., Khallouf, K.A., Gibson, R., Hajianpour, A. & Mathew, C.G. A new autosomal recessive anomaly mimicking Fanconi's anaemia phenotype. *Arch Dis Child* **68,** 101-3 (1993).
23. Germanaud, D. *et al.* Simplified gyral pattern in severe developmental microcephalies? New insights from allometric modeling for spatial and spectral analysis of gyrification. *Neuroimage* **102 Pt 2**, 317-31 (2014).
24. Martin, C.A. *et al.* Mutations in PLK4, encoding a master regulator of centriole biogenesis, cause microcephaly, growth failure and retinopathy. *Nat Genet* **46**, 1283-92 (2014).
25. Trimborn, M. *et al.* Mutations in microcephalin cause aberrant regulation of chromosome condensation. *Am J Hum Genet* **75**, 261-6 (2004).
26. Griffith, E. *et al.* Mutations in pericentrin cause Seckel syndrome with defective ATR-dependent DNA damage signaling. *Nat Genet* **40**, 232-6 (2008).
27. Yeo, G. & Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**, 377-94 (2004).
28. Mouse Genome Database (MGD) at the Mouse Genome Informatics website. Vol. 2016 (ed. Laboratory, T.J.) (Bar Harbor, Maine).
29. Bandura, J.L. *et al.* humpty dumpty is required for developmental DNA amplification and cell proliferation in Drosophila. *Curr Biol* **15**, 755-9 (2005).
30. Cortez, D. Preventing replication fork collapse to maintain genome integrity. *DNA Repair (Amst)* **32**, 149-57 (2015).
31. Errico, A. & Costanzo, V. Mechanisms of replication fork protection: a safeguard for genome stability. Crit Rev *Biochem Mol Biol* **47**, 222-35 (2012).
32. Fuchs, F. *et al.* Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. *Mol Syst Biol* **6**, 370 (2010).

565 33. Papadopoulos, D.K. *et al.* Probing the kinetic landscape of Hox transcription factor-
566 DNA binding in live cells by massively parallel Fluorescence Correlation
567 Spectroscopy. *Mech Dev* **138 Pt 2**, 218-25 (2015).
568 34. Bacia, K. & Schwille, P. Practical guidelines for dual-color fluorescence cross-
569 correlation spectroscopy. *Nat Protoc* **2**, 2842-56 (2007).
570 35. Sirbu, B.M. *et al.* Analysis of protein dynamics at active, stalled, and collapsed
571 replication forks. *Genes Dev* **25**, 1320-7 (2011).
572 36. Liu, S. *et al.* ATR autophosphorylation as a molecular switch for checkpoint
573 activation. *Mol Cell* **43**, 192-202 (2011).
574 37. Durkin, S.G., Arlt, M.F., Howlett, N.G. & Glover, T.W. Depletion of CHK1, but not
575 CHK2, induces chromosomal instability and breaks at common fragile sites.
576 *Oncogene* **25**, 4381-8 (2006).
577 38. Ozeri-Galai, E., Schwartz, M., Rahat, A. & Kerem, B. Interplay between ATM and
578 ATR in the regulation of common fragile site stability. *Oncogene* **27**, 2109-17 (2008).
579 39. Toledo, L.I. *et al.* ATR prohibits replication catastrophe by preventing global
580 exhaustion of RPA. *Cell* **155**, 1088-103 (2013).
581 40. Brown, E.J. & Baltimore, D. ATR disruption leads to chromosomal fragmentation
582 and early embryonic lethality. *Genes Dev* **14**, 397-402 (2000).
583 41. Brown, E.J. & Baltimore, D. Essential and dispensable roles of ATR in cell cycle
584 arrest and genome maintenance. *Genes Dev* **17**, 615-28 (2003).
585 42. Forment, J.V., Blasius, M., Guerini, I. & Jackson, S.P. Structure-specific DNA
586 endonuclease Mus81/Eme1 generates DNA damage caused by Chk1 inactivation.
587 *PLoS One* **6**, e23517 (2011).
588 43. Couch, F.B. *et al.* ATR phosphorylates SMARCAL1 to prevent replication fork
589 collapse. *Genes Dev* **27**, 1610-23 (2013).

590 44. Ragland, R.L. *et al.* RNF4 and PLK1 are required for replication fork collapse in
591 ATR-deficient cells. *Genes Dev* **27**, 2259-73 (2013).
592 45. Hodskinson, M.R. *et al.* Mouse SLX4 is a tumor suppressor that stimulates the
593 activity of the nuclease XPF-ERCC1 in DNA crosslink repair. *Mol Cell* **54**, 472-84
594 (2014).
595 46. Svendsen, J.M. *et al.* Mammalian BTBD12/SLX4 assembles a Holliday junction
596 resolvase and is required for DNA repair. *Cell* **138**, 63-77 (2009).
597 47. Takahashi, T., Nowakowski, R.S. & Caviness, V.S., Jr. The cell cycle of the
598 pseudostratified ventricular epithelium of the embryonic murine cerebral wall. *J*
599 *Neurosc*i **15**, 6046-57 (1995).
600 48. Snow, M.H.L. Gastrulation in the mouse: growth and regionalization of epiblast. *J.*
601 *Embryol. Exp. Morphol* **42**, 293–303 (1977).
602 49. Murga, M. *et al.* A mouse model of ATR-Seckel shows embryonic replicative stress
603 and accelerated aging. *Nat Genet* **41**, 891-8 (2009).
604 50. Despras, E., Daboussi, F., Hyrien, O., Marheineke, K. & Kannouche, P.L.
605 ATR/Chk1 pathway is essential for resumption of DNA synthesis and cell survival in
606 UV-irradiated XP variant cells. *Hum Mol Genet* **19**, 1690-701 (2010).
607 51. Kawabata, T. *et al.* Stalled fork rescue via dormant replication origins in
608 unchallenged S phase promotes proper chromosome segregation and tumor
609 suppression. *Mol Cell* **41**, 543-53 (2011).
610 52. Kumagai, A., Lee, J., Yoo, H.Y. & Dunphy, W.G. TopBP1 activates the ATR-ATRIP
611 complex. **Cell** *124*, 943-55 (2006).
612 53. Bass, T.E. *et al.* ETAA1 acts at stalled replication forks to maintain genome integrity.
613 *Nat Cell Biol* **18**, 1185-1195 (2016).
614 54. Haahr, P. *et al.* Activation of the ATR kinase by the RPA-binding protein ETAA1. *Nat*
615 *Cell Biol* **18**, 1196-1207 (2016).

55. Duursma, A.M., Driscoll, R., Elias, J.E. & Cimprich, K.A. A role for the MRN complex in ATR activation via TOPBP1 recruitment. *Mol Cell* **50**, 116-22 (2013).

**Figure Legends:**

**Figure 1: *DONSON* mutations cause severe microcephaly and short stature.**

**(a)** *DONSON* mutations result in severe prenatal-onset microcephaly, often associated with short stature. Length at birth (Lgt), current height (Hgt) and head circumference (OFC) plotted as z-scores (SD from population mean for age and sex). Black horizontal bars indicate mean values. Dashed line at -2 SD indicates cut-off for normal population distribution. Patients from the three independently identified *DONSON* patient cohorts are denoted by black (P1-P12 and P20), orange (P13), and blue (P14-19 and P21) circles. **(b)** Photographs of affected individuals with *DONSON* mutations demonstrating facial similarities. Written consent to publish photographs was obtained from the affected families. P, patient. **(c)** Schematics of the *DONSON* gene and protein indicating position of the identified mutations. DONSON mutations comprised a range of mutation classes (nonsense, frameshift, essential splice site, missense and intronic). The genomic structure is based on the longest ORF containing ten coding exons (white rectangles) (NM_017613.3). The positions of identified mutations affecting splicing are shown on the gene structure (top) and missense and truncating variants on the encoded protein (bottom).

**Figure 2: Mutations in *DONSON* affect protein levels.**

**(a-b)** *DONSON* mutations result in severely reduced levels of DONSON protein. Immunoblotting of cell extracts from lymphoblastoid (a) and fibroblast (b) cell lines derived from patients with mutations in *DONSON*. ATR was used as a loading control. The two blots from (a) originate from two independent gels. **(c)** The K489T, but not the S28R variant, associated with the *DONSON* haplotype affects protein levels. Cells were treated with doxycycline 48 h post siRNA transfection, and harvested for Western blot analysis 24 h later (n=2). Exogenous DONSON were detected using an anti-GFP antibody

643 respectively. TOPBP1 was used as a loading control. Depletion of endogenous DONSON
644 in these cells was confirmed by immunoblotting (**Supplementary Fig. 2**).
645
646 **Figure 3. Loss of DONSON results in spontaneous replication fork stalling and**
647 **increased genome instability.**
648 **(a)** DONSON protein levels are increased during S-phase. HeLa cells were synchronised
649 in S-phase using a double thymidine block, released, harvested at the indicated time
650 points, and immunoblotting was performed (n=2). Cyclin A and phospho-histone H3 Ser-
651 10 are markers of S/G2 and M phase respectively. Vinculin represents a loading control.
652 **(b)** S-phase is prolonged upon DONSON depletion. HeLa cells transfected with the
653 indicated siRNAs were pulsed with BrdU, fixed and analysed by FACS (n=4; error bars
654 indicate SD). **(c-e)** Replication fork analysis of HeLa cells transfected with control or
655 DONSON siRNA and pulsed with CldU and IdU. **(c)** Top: Schematic of DNA fibre analysis.
656 Bottom: loss of DONSON does not decrease replication fork velocity. Replication fork
657 speed (kb/min) was determined (n=5). **(d)** DONSON depletion results in spontaneous fork
658 stalling. Percentages of ongoing replication forks, new origins and stalled replication forks
659 in cells from (c) were quantified (n=3). **(e)** DONSON depletion leads to replication fork
660 asymmetry. Top: example images; magenta arrows indicate origins of replication; white
661 arrow denotes fork asymmetry. Bottom: plot indicates the ratio of left/right fork track
662 lengths of bidirectional replication forks in cells from (c). Red lines denote median ratios
663 (n=3). **(f)** Loss of DONSON increases spontaneous γH2AX/53BP1 foci formation. HeLa
664 cells transfected with the indicated siRNAs were immunostained with antibodies to 53BP1
665 and γH2AX (left panel), and the percentage of cells with >10 53BP1 and γH2AX foci were
666 quantified using fluorescence microscopy (right panel; n=5; >300 cells per sample per
667 independent experiment). Scale bar; 10 μm.
668

**Figure 4. DONSON localizes to the replication fork.**

**(a-d)** DONSON interacts with multiple components of the replication machinery. **(a)** GFP or GFP-DONSON was precipitated by GFP-Trap, from asynchronous cells or cells accumulated in S-phase with 2 mM HU treatment for 24 h. Heatmap denotes significant interactions identified by mass spectrometry (n=3). Inset: Schematic of the mammalian replisome with selected replication factors. **(b)** 293FT cells were transfected with the indicated expression vectors in the presence/absence of HU. GFP or GFP-DONSON were isolated by GFP-Trap and co-precipitating proteins visualised by immunoblotting (n=2). Benzonase Nuclease was included to exclude DNA-mediated interactions. The bottom two panels are scanned images of Ponceau S-stained nitrocellulose membrane. **(c-d)** DONSON localises in close proximity to replication forks. **(c-d)** PLA was carried out on cells from (a) using the indicated antibodies in the presence/absence of HU (n=2). **(c)** Quantification of PLA signals. **(d)** Representative PLA images. **(e-f)** DONSON interacts with PCNA at replication foci in live cells. **(e)** Representative confocal images of live cells expressing GFP-DONSON and RFP-PCNA. Boxes indicate representative regions used for FCCS analysis. **(f)** FCCS measurements of GFP-DONSON and RFP-PCNA reveal significant cross-correlation at replication foci at similar concentrations. Average cross-correlation curves are shown from cells expressing GFP-DONSON in replication foci (red) or non-replicating (grey) cells, or GFP-expressing S-phase nuclei (purple). Inset: Mean cross-correlation amplitude values from multiple cells (error bars indicate SD; n=4, 3 and 5). Increased G(τ) values indicate higher degree of cross-correlation between GFP-DONSON and RFP-PCNA in replication foci. See also **Supplementary Fig. 11**. **(g)** iPOND was performed on 293T (n=3), HeLa (n=2) and HCT116 (n=2) cells, and EdU-coprecipitates analysed by mass spectrometry. Data represents the combination of all seven experiments. Log2 abundance denotes the ratio of proteins at nascent DNA

compared to mature chromatin. Values >0 represent proteins enriched at forks, whilst values ≤0 denote chromatin-bound factors. Scale bars; 10 μm.

**Figure 5. Depletion of DONSON compromises activation of cell cycle checkpoints.**

**(a-c)** Loss of DONSON results in replication fork instability that is exacerbated by replication stress. **(a)** HeLa cells transfected with either control or DONSON siRNA were pulsed with CldU, exposed to 2 mM HU for 2 h, and then pulsed with IdU. Alternatively, cells were exposed to 50 ng/ml MMC for 24 h, and pulsed with sequential pulses of CldU and IdU (see schematic). DNA fibres were quantified, and the percentage of **(b)** stalled forks and **(c)** new origins are displayed (in all cases n=3). **(d)** Loss of DONSON is epistatic with ATR inhibition. Replication fork analysis of HeLa cells transfected with either control or DONSON siRNA. Cells were pulsed with CldU, exposed to 2 mM HU +/- 5μM ATR inhibitor for 2 h, and then pulsed with IdU (n=3). New origins ($2^{nd}$ label origin) were counted as an indicator of intra-S phase checkpoint activation. **(e)** Cells lacking DONSON exhibit defective or delayed ATR activation in response to replication stress. Whole cell extracts of HeLa cells transfected with either control or DONSON siRNA were subjected to immunoblot analysis using the indicated antibodies following treatment with 1 mM HU (n=2). **(f)** The percentage of mitotic cells following exposure to 1 mM HU for 24 h (from (e)) was determined by flow cytometry, using antibodies to phosphorylated histone H3-Ser10 (a marker of mitosis) (n=5).

**Figure 6. Increased spontaneous chromosome breakage and fragmentation of mitotic chromosomes in DONSON-depleted cells.**

**(a,b)** Metaphases chromosomes from DONSON or control siRNA transfected HeLa cells were visualised by Giemsa staining and light microscopy. **(a)** Quantification of average numbers of chromatid gaps/breaks per metaphase (n=6; >50 metaphases per sample per

720 experiment). **(b)** Representative images of normal chromosomes, chromosomes
721 containing gaps/breaks, highly fragmented and pulverized chromosomes. Red arrows
722 denote chromatid gaps/breaks; blue arrows indicate chromosomal exchanges. Scale bar;
723 10 µm. **(c-g)** Loss of the structure-specific nucleases MUS81 or XPF significantly reduces
724 the spontaneous replication fork asymmetry and genome instability in DONSON-depleted
725 cells. **(c)** Cells transfected with the indicated siRNAs were pulsed with CldU and IdU.
726 Replication fork asymmetry was measured as in (**Fig. 3e**). The red lines denotes median
727 ratios (n=3). **(d)** Co-depletion of MUS81 or XPF with DONSON reduces levels of
728 spontaneous DNA damage. Extracts from cells transfected with the indicated siRNAs were
729 subjected to SDS-PAGE and immunoblotting using the antibodies indicated. **(e-f)** Co-
730 depletion of MUS81 (e) or XPF (f) reduces chromosomal aberrations in cells lacking
731 DONSON. Quantification of the average number of chromatid gaps/breaks per metaphase
732 in cells transfected with control, DONSON, MUS81 and/or XPF siRNA. At least 50
733 metaphases per experiment were counted (n=3). **(g)** Quantification of the average
734 percentage of metaphases containing highly fragmented chromosomes or pulverized
735 chromosomes in cells transfected with the indicated siRNAs. At least 50 metaphases per
736 experiment were counted (n=3).
737
738 **Figure 7: DONSON patient cells have spontaneous defects in replication fork**
739 **progression that result in DNA damage**
740 **(a)** Complementation of patient-derived fibroblasts with WT *DONSON*. Fibroblasts derived
741 from DONSON patients P2, P6 and P9 were infected with retroviruses encoding either WT
742 *DONSON* or an empty vector. DONSON expression was determined by immunoblotting. A
743 non-specific cross-reactive protein represents a loading control. **(b)** Expression of WT
744 *DONSON* in patient fibroblasts rescues elevated levels of spontaneous DNA damage. The
745 percentage of cells from (a) with 53BP1/γH2AX foci was quantified by immunostaining

(n=3). **(c)** DNA fibre analysis of complemented DONSON patient fibroblasts pulsed with CldU and IdU. Fork asymmetry was quantified. Plot indicates ratios of left/right fork track lengths of bidirectional replication forks. The red lines denote median ratios. (n=3). **(d)** The percentage of stalled forks and new origins from cells in (c) was quantified (n=3). Ongoing forks are shown in (**Supplementary Fig. 19**).

**Table 1: Biallelic *DONSON* mutations identified in 29 individuals**

| Patient | Country of Origin | Mutation 1 | Mutation 2 | Segregation |
|---|---|---|---|---|
| P1-1 | Italy | c.1251_1256delCTCTAA, p.Asn417_Ser418del | *haplotype* | Het, M |
| P1-2 | Italy | c.1251_1256delCTCTAA, p.Asn417_Ser418del | *haplotype* | Het, M |
| P2 | UK | c.877C>T, p.Arg293* | *haplotype* | Het, M, P |
| P3 | UK | c.1254dupT, p.Lys419* | *haplotype* | Het, M, P |
| P4 | UK | c.1686dupT, p.Asn563* | *haplotype* | Het, nps |
| P5 | Somalia | c.832T>C, p.Cys278Arg AND/OR c.845A>G, p.Tyr282Cys | *haplotype* | Het, M, P |
| P6 | USA | c.1282C>T, p.Gln428* | *haplotype* | Het, M, P |
| P7 | USA | c.1282C>T, p.Gln428* | *haplotype* | Het, nps |
| P8 | Italy | c.1474_1475delCA, p.Gln492Glufs*18 | c.786-7T>C | Het, M, P |
| P9 | Turkey | c.876C>G, p.Phe292Leu | c.1047-9A>G (SS) | Het, M |
| P10-1 | India | c.1628_1630dupAAA, p.Gln543_Ile544insLys | c.1032C>T, p.Ser344Ser (SS) | Het, M, P |
| P10-2 | India | c.1628_1630dupAAA, p.Gln543_Ile544insLys | c.1032C>T, p.Ser344Ser (SS) | Het, M, P |
| P11 | Saudi Arabia | c.786-22A>G (SS) | c.786-22A>G (SS) | Hom, M, P |
| P12 | Saudi Arabia | c.786-22A>G (SS) | c.786-22A>G (SS) | Hom, nps |
| P13-1 | Palestine | c.1337T>C, p.Met446Thr | c.1337T>C, p.Met446Thr | Hom, M, P |
| P13-2 | Palestine | c.1337T>C, p.Met446Thr | c.1337T>C, p.Met446Thr | Hom, M, P |
| P13-3 | Palestine | c.1337T>C, p.Met446Thr | c.1337T>C, p.Met446Thr | Hom, M, P |
| P14 | Saudi Arabia | c.786-22A>G (SS) | c.786-22A>G (SS) | Hom, M, P |
| P15 | Saudi Arabia | c.786-22A>G (SS) | c.786-22A>G (SS) | Hom, M, P |
| P16 | Saudi Arabia | c.786-22A>G (SS) | c.786-22A>G (SS) | Hom, M, P |
| P17 | Saudi Arabia | c.786-22A>G (SS) | c.786-22A>G (SS) | Hom, M, P |
| P18-1 | Saudi Arabia | c.786-22A>G (SS) | c.786-22A>G (SS) | Hom, M, P |
| P18-2 | Saudi Arabia | c.786-22A>G (SS) | c.786-22A>G (SS) | Hom, M, P |
| P18-3 | Saudi Arabia | c.786-22A>G (SS) | c.786-22A>G (SS) | Hom, M, P |
| P19 | Turkey | c.1297C>T, p.Pro433Ser | c.1297C>T, p.Pro433Ser | Hom, M, P |
| P20-1 | South Africa | c.1254dupT, p.Lys419* | c.1510G>A, p.Glu504Lys | Het, M, P |
| P20-2 | South Africa | c.1254dupT, p.Lys419* | c.1510G>A, p.Glu504Lys | Het, M, P |
| P21-1 | Saudi Arabia | c.1047-9A>G (SS) | c.1047-9A>G (SS) | Hom, M, P |
| P21-2 | Saudi Arabia | c.1047-9A>G (SS) | c.1047-9A>G (SS) | Hom, M, P |

'Haplotype' indicates the presence of three co-segregating variants: c.82A>C (p.Ser28Arg); c.786-33A>G; c.1466A>C (p.Lys489Thr). Hom, homozygous in affected individual; Het, compound heterozygous in affected individual; M, mutation identified in mother; P, mutation identified in father; nps, no parental samples available; SS, Splice site mutation. Reference sequence, NM_017613.3.

## Online Methods

### Research subjects.

Genomic DNA from the affected children and family members was extracted from peripheral blood using standard methods or saliva samples using Oragene collection kits according to the manufacturer's instructions. Informed consent was obtained from all participating families. Ethics for the studies were approved by the Scottish Multicentre Research Ethics Committee (04:MRE00/19), by an IRB-approved research protocol (KFSHRC RAC# 2080006), and via the 'National Gene Mapping' protocol by Guy's and St. Thomas' National Health Service (NHS) Foundation Trust local research ethics committee (ref.: 08/H0802/84, "Systematic Characterization of Genes in Inherited Disorders"). In addition, ethical approval for linkage studies on the genetics of Fanconi anaemia in 1989 were obtained from the Guy's Hospital Research Ethics Committee (ref. EC89/10/27)[22], with further approval for mutation analysis on existing samples in 1996 (ref. 96/3/9). Parents provided written consent for the publication of photographs of the affected individuals.

### Exome sequencing and haplotype analysis.

Exome sequencing of genomic DNA and variant filtering was performed as described previously[12]. Cohort resequencing was performed by Sanger sequencing of PCR products representing all coding exons of *DONSON* (primer sequences are detailed in **Supplementary Table 4**), with variant calling using MutationSurveyor (SoftGenetics Inc.). Haplotype analysis was undertaken by SNP genotyping both patients using Affymetrix CytoScan 750K arrays. Genotypes were generated using Affymetrix Genotyping Console software and examined manually. The pathogenic impact of *DONSON* mutations was predicted using Alamut Visual Software (Interactive Biosoftware Inc).

**Cell culture and generation of cell lines.**

Lymphoblastoid cell lines (LCLs) were maintained in RPMI 1640 supplemented with 15% FBS, L-glutamine and penicillin/streptomycin antibiotics. LCLs were generated in house from peripheral blood samples by EBV transformation using standard methods. Dermal primary fibroblasts were grown from skin-punch biopsies in AmnioMax medium (Life Technologies) and then maintained in DMEM supplemented with 10% FBS, 5% L-glutamine and 5% penicillin/streptomycin antibiotics. Patient cell lines were validated using Sanger sequencing and immunoblotting. 293FT (Invitrogen) and HeLa (ATCC) cells were maintained in DMEM supplemented with 10% FBS, 5% L-glutamine and 5% penicillin/streptomycin antibiotics.

Stable cell lines were generated by Flp recombinase–mediated integration using HeLa-Flp-In T-REx host cells (gift from S. Taylor, University of Manchester) transfected with pcDNA5/FRT/TO-EGFP (vector only or EGFP-TRAIP) and pCAGGS-Flp.e (gift from D.-J. Kleinjan, University of Edinburgh). Transfected cells were selected using 5 µg/ml blasticidin and 400 µg/ml hygromycin, and the resulting colonies were then expanded for testing. Protein expression was induced with 1 µg/ml doxycycline (Sigma-Aldrich) treatment.

Primary fibroblasts derived from patients 2, 6 and 9 were immortalized with *TERT* retroviral supernatant with 4 µg/ml polybrene and infected with pMSCV-vector only or pMSCV-*DONSON*. Selection was performed using 750 ng/ml puromycin (Clontech) and 500 µg/ml neomycin (Invitrogen). Expression of the protein was verified by immunoblotting (**Fig. 7a**). All cell lines were routinely tested for mycoplasma.

**Cell treatments.**

Plasmids and siRNA oligos were transfected in Opti-MEM reduced serum medium using Lipofectamine 2000 and Oligofectamine (Life Technologies) respectively according to the manufacturer's guidelines. A custom siRNA targeting lacZ has previously been described[56], and was used as a control siRNA. A DONSON siRNA SMARTpool (Dharmacon) was used for all siRNA transfections except when transfecting the HeLa Flp-In/T-Rex cells expressing an exogenous, siRNA-resistant, GFP-tagged DONSON construct. In this case a custom DONSON siRNA sequence (CCTGTGGACTGGAGTATTAdTdT) was used (Dharmacon). MUS81 siRNA SMARTpool and XPF siRNA SMARTpool (Dharmacon) were used where indicated. Transfected cells in both cases were analysed at 48-72 h post transfection. Where indicated, cells were treated with 1-2 mM hydroxyurea (Sigma-Aldrich), 50 ng/ml mitomycin C (Sigma-Aldrich) or 2 mM thymidine (Sigma-Aldrich). The ATR inhibitor (VE-821; Selleck Chemicals) was used at 5 μM. dNTP analogues EdU, CldU and IdU were purchased from Sigma Aldrich, and were used as indicated.

**RT-PCR.**

Total RNA was extracted from cell lines using the RNeasy kit (Qiagen) according to the manufacturer's instructions. DNA was removed by treatment with DNase I (Qiagen), and cDNA was generated using random oligomer primers and AMV RT (Roche). The RT-PCR primer pairs used are detailed in **Supplementary Table 4**.

**DNA expression constructs**

pEGFP-DONSON expression construct was created by cloning the human *DONSON* ORF into the pDONR221 Gateway shuttle vector (Invitrogen). WT *DONSON* was amplified from cDNA and recombined into the pDEST-EGFP vector to generated a GFP-tagged DONSON expression construct. The *DONSON* ORF was made siRNA resistant using site-

836 directed mutagenesis (Agilent Technologies) by altering the following nucleotides:
837 CCTGTGGACTGGAGTATTA was changed to CCCGTAGATTGGTCTATCA. Patient-
838 associated mutations were engineered into the pEGFP-DONSON expression plasmid
839 using site-directed mutagenesis according to the manufacturer's instructions. (All primers
840 are detailed in **Supplementary Table 4**).
841
842 The retroviral expression construct expressing DONSON was created by recombination
843 between the pDONR221-DONSON vector and a Gateway-compatible pMSCVneo
844 retroviral expression construct (Clontech).
845
846 The human telomerase reverse transcriptase (hTERT) expressing retroviral construct used
847 to immortalise patient-derived human fibroblasts was a kind gift from Bob Weinberg
848 (Addgene plasmid: #1771).
849
850 **Minigene splicing reporter assay**
851 A 1.58 kb stretch of the *DONSON* gene encompassing the 3' end of intron 3, exon 4,
852 intron 4, exon 5 and the 5' end of intron 5 was amplified using DNA from a healthy
853 individual and *DONSON* patients (carrying mutation c.786-22A>G or c.786-33A>G) using
854 the DONSON-int3-SalI-F and DONSON-int5-SpeI-R primers, and cloned into the RHCglo
855 vector[57] using the SalI and SpeI restriction sites. Site-directed mutagenesis was used to
856 introduce the DONSON intron 4 splice acceptor mutation (c.786-1G>A) into the splicing
857 reporter construct. HeLa cells were transfected with each individual splicing mutation
858 reporter construct using Lipofectamine 2000 according the manufacturer's instructions. 24
859 h post-transfection, cells were harvested, total cellular RNA was extracted and cDNA
860 generated using Superscript III reverse transcriptase first-stand synthesis system
861 (Invitrogen). PCR was carried out using primers (RSV_minigene_F and RSV_minigene_R)

to the 5' and 3' ends of the artificial exons present in the RHCglo vector. *DONSON* WT and mutant cDNA amplicons were resolved on a 2% agarose gel to visualise differences in splicing. Individual PCR products were subsequently cloned into the pGEM-T Easy Vector (Promega) and sequenced to verify the exon content of each transcript. All relevant primers are detailed in **Supplementary Table 4**.

**iPOND**

iPOND was performed as previously described[35,58]. Briefly, exponentially growing cells were incubated with 10 µM EdU for 10 min, cross-linked with 1% formaldehyde, harvested and permeabilised. For pulse-chase controls, cells were incubated in 10 µM EdU for 10 min, washed in media containing 10 µM thymidine, then incubated with media containing 10 µM thymidine for 1 h, before being cross-linked. Biotin azide was covalently attached to EdU within newly replicated DNA using a Click reaction, and EdU containing DNA was precipitated using Streptavidin agarose beads. Edu co-precipitates were then analysed by mass spectrometry. Log2 abundance values represent the ratio of proteins found in EdU-pulsed samples compared to those pulse-chased with EdU-thymidine.

**Immunoblot analysis and antibodies**

Whole cell extracts were obtained by sonication in UTB buffer (8 M Urea, 50 mM Tris, 150 mM β-mercaptoethanol) and analysed by SDS-PAGE following standard procedures. Protein samples were run on 6-12% acrylamide SDS-PAGE or 4-12% NuPage mini-gels (Life Technologies) and transferred onto nitrocellulose membrane. Immunoblotting was performed using antibodies to: Cyclin A (Santa Cruz, sc-751; 1:1,000), CHK1 (Santa Cruz, sc-8408; 1:1,000), CHK2 (Santa Cruz, sc-5278; 1:1000), FANCD2 (Santa Cruz, sc-20022; 1:1000), MCM2 (BD Transduction Laboratories, 610700; 1:10000), MCM7 (Santa Cruz, sc-56324; 1:1000), MUS81 (Santa Cruz, sc-53382; 1:2000); XPF (Santa Cruz, sc-136153;

1:1000); H2A (Millipore, 07-146; 1:3000), γ-H2AX (Millipore, 05-636; 1:3000), RPA2 (Millipore, NA18; 1:1000), phospho-histone H3 Ser-10-P (Millipore); pS343-NBS1 (Abcam, 47272; 1:500); NBS1 (Genetex, GTX70224; 1:10000); ATR (Bethyl Laboratories, A300-137A; 1:1000), pS345-CHK1 (Cell Signaling Technology, 2341; 1:100), pS4/S8-RPA2 (Bethyl Laboratories, A300-245A; 1:1,000), pS966-SMC1 (Bethyl Laboratories, A300-050A; 1:1,000), SMC1 (Bethyl Laboratories, A300-055A; 1:1,000), Treslin (Bethyl Laboratories, A303-472A; 1:1,000); TOPBP1 (Bethyl Laboratories; A300-111A; 1:1000); Vinculin (Sigma-Aldrich, V9264; 1:1,000); α-Tubulin (Sigma-Aldrich, T5168; 1:4000); GFP (Roche, 11814460001; 1:500). The polyclonal anti-DONSON antibody was generated by immunising rabbits with a GST-fusion protein encoding aa 1-125 of human DONSON. Antibody was affinity-purified from rabbit sera (Eurogentec) and specificity established using lysates from patient cells and DONSON siRNA-transfected cells.

Loading controls for all blots derive from reprobing the same membrane, except for phospho-antibody blots, where paired gels were run simultaneously, and blotted in parallel for phosphorylated and total proteins.

**Immunofluorescence and fluorescent microscopy.**

siRNA transfected HeLa cells or passage-matched *TERT*-immortalized fibroblasts were seeded on coverslips 24 h before extraction/fixation. To visualise cells undergoing DNA replication, cells incubated in medium containing 10 μM EdU for 10-30 min prior to harvesting. To remove soluble proteins before immunofluorescence, cells were pre-extracted for 10 min on ice with ice-cold buffer (25 mM HEPES, pH 7.4, 50 mM NaCl, 1 mM EDTA, 3 mM MgCl2, 300 mM sucrose and 0.5% Triton X-100) and then fixed with 4% paraformaldehyde for 15 min. For analysis of cells transfected with GFP-tagged protein, cells were fixed and permeabilised by incubation with ice-cold methanol for 20 minutes.

EdU immunolabeling was performed using Click-iT EdU Imaging Kit (Invitrogen, C10337) according to the manufacturer's protocol. Cells were stained for 53BP1 (Novus Biologicals, NB100-304; 1:1,000) and/or γH2AX (Millipore, 05-636; 1:1000) and stained with secondary antibodies conjugated to Alexa Fluor-568 (Life Technologies) and DAPI.

For quantification of signal-integrated densities of γH2AX staining, cells were stained with an antibody specific to γH2AX (Millipore, 05-636; 1:1000), images were visualized using a Zeiss Axioplan 2 microscope with iVision software (BioVision Technologies) and captured using a 40× oil-immersion objective. For quantification of signal-integrated densities of RPA staining, cells were stained with RPA2 antibody (Millipore, NA18; 1:200), images were visualized using a Nikon Eclipse Ni microscope with NIS-Elements software (Nikon Instruments) and captured using a 100× oil-immersion objective. Nuclei were segmented on the basis of DAPI staining and then signal-integrated density of γH2AX or RPA staining quantified for each nuclear region using ImageJ software (US National Institutes of Health). For quantification of γH2AX staining, more than 100 EdU positive cells and 50 EdU negative cells were analyzed per experiment per condition, and for quantification of RPA staining, more than 200 cells were analyzed per experiment per condition. Exposure time, binning, microscope settings and light source intensity were kept constant for all the samples in all cases.

For quantification of native BrdU foci cells were incubated in medium containing 10 μM BrdU for 24 h prior to harvesting. Six hours prior to harvesting, 2 mM HU was added to the media. To visualise ssDNA foci, cells were extracted for 10 min on ice with ice-cold buffer (25 mM HEPES, pH 7.4, 50 mM NaCl, 1 mM EDTA, 3 mM MgCl2, 300 mM sucrose and 0.5% Triton X-100) and then fixed with 4% paraformaldehyde for 15 min. After fixation,

cells were washed with PBS and blocked in 3% FCS in PBS for 30 min at room temperature. ssDNA was visualised using a BrdU antibody (Abcam, ab6326; 1:500). To denature DNA cells were incubated in 2 M HCl in PBS for 30 min prior to addition of the BrdU antibody. Images were acquired as for γH2AX quantification and foci were quantified using ImageJ-based script. Nuclei were defined on the basis of DAPI staining and native BrdU foci were detected using "Find maxima" function of ImageJ within each nuclear region. Exposure time, binning, microscope settings, light source intensity and the noise level in “Find maxima” function were kept constant for all the samples within each individual experiment. More than 100 cells were analyzed per experiment per condition.

**Metaphase spreads**

Chromosomal aberrations were scored in Giemsa-stained metaphase spreads as previously described[56]. Briefly, demecolcine (Sigma Aldrich) was added at a final concentration of 0.2 µg/ml 4 h prior to harvesting. Cells were harvested by trypsinisation, subjected to hypotonic shock for 1 hour at 37°C in 0.3 M sodium citrate and fixed in 3:1 methanol:acetic acid solution. Cells were dropped onto acetic acid humidified slides, stained for 15 minutes in Giemsa-modified solution (Sigma; 5% v/v in $H_2O$) and washed in water for 5 minutes.

**DNA fibre spreading assay**

Passage-matched primary, *TERT*-immortalized fibroblasts or siRNA transfected HeLa cells were pulse labeled with CldU for 20 min, washed with media and damaged with 2 mM hydroxyurea for 2 h before being pulse labeled with IdU for 40 min. Alternatively, 50 ng/ml mitomycin C was added to the cells 24 h before CldU pulse labeling and left on during 20 min CldU and 20 min IdU pulse labeling. Cells were harvested by trypsinization, and cell pellets were washed in PBS. $5 \times 10^5$ cells were lysed directly onto glass slides using

spreading buffer (200 mM Tris-HCl, pH 7.5, 50 mM EDTA, 0.5% SDS) and fixed in methanol/acetic acid (3:1 ratio). Following 2.5 M HCl denaturation, CldU was detected using rat anti-BrdU (clone BU1/75, ICR1; Abcam, ab6326; 1:750), and IdU was detected using mouse anti-BrdU (clone B44; BD Biosciences, 347583; 1:750). Slides were then fixed in 4% paraformaldehyde before immunostaining with secondary antibodies conjugated to Alexa Fluor-594 or Alexa Fluor-488 (Life Technologies). Labeled DNA fibers were visualized using a Nikon Eclipse Ni microscope with NIS-Elements software (Nikon Instruments). Images were captured using 40× oil-immersion objectives and were recoloured and analyzed using ImageJ software (US National Institutes of Health). For quantification of replication structures, at least 400 structures were counted per experiment. Tract lengths were measured using ImageJ (National Institutes of Health; http://rsbweb.nih.gov/ij/). To calculate fork velocity, arbitrary length values were converted into micrometers using the scale bars created by the microscope, with 1 μm equivalent to 2.59 kb[59]. Replication fork speed (kb/min) was then determined by dividing the length of CldU and IdU tracks (in kb) from ongoing forks by the pulse time.

**FACS analysis**

For BrdU analysis, HeLa cells were pulse labeled with 10 μM BrdU for 30 min before fixation with 70% ethanol at −20 °C for 16 h. Cells were then digested with 1 mg/ml pepsin and denatured with 2 M HCl before washing with PBS and blocking in 0.5 % BSA, 0.5 % Tween-20. BrdU labeling was detected using anti-BrdU antibody (Abcam, ab6326; 1:75) and FITC-conjugated anti-rat secondary antibody. DNA content was assessed by staining with 50 μg/ml propidium iodide. Cells were sorted on a BD Biosciences FACS Aria II and data were analyzed using FlowJo software (v7.6.1, Tree Star).

For mitotic analysis and immuno-detection of phospho-histone H3 (Ser10), HeLa cells were harvested, fixed, permeabilised 24 h post exposure to HU or MMC, as previously described[56]. Cells were analysed using an Accuri flow cytometer (BDBiosciences) in conjunction with CFlowplus software. Data represents that obtained from at least 30,000 cells.

**Immunoprecipitation and GFP-Trap**

293FT cells transfected with plasmids encoding GFP-DONSON or GFP were untreated, or exposed to 2 mM HU for 16 h and harvested. Cells were then incubated in lysis buffer (150 mM NaCl, 50 mM Tris HCl pH7.5, 2 mM $MgCl_2$, 1 % NP40, 90 U/ml Benzonase (Novagen) and Protease Inhibitor Cocktail EDTA free (Roche)) for 30 min with rotation at 4 °C. The resultant cell lysates were pre-cleared at 44,000 rpm at 4 °C for 30 min.

For immunoprecipitations, 3 mg of lysate was immunoprecipitated with 5 µg of antibody, immune complexes collected with Protein A-Sepharose (Sigma-Aldrich). Complexes were washed with wash buffer (150 mM NaCl, 50 mM Tris HCl pH 7.5, 0.5 % NP40, and Complete Protease Inhibitor Cocktail (Roche)) and analysed by SDS-PAGE.

For GFP-Trap, 3 mg lysates were incubated with GFP-Trap agarose beads (ChromoTek) at 4 °C for 5 h. The resulting GFP-Trap complexes were washed with wash buffer as above and analysed by SDS-PAGE. Experiments were carried out in the presence of Benzonase Nuclease to exclude the possibility of interactions being mediated by DNA.

For mass spectrometry analysis, GFP or GFP-DONSON were isolated from tetracycline induced, or uninduced, Flp-In T-REx HeLa cell extracts by incubation with GFP-trap magnetic agarose beads (Chromotek) for 2 hours on a Kingfisher Duo robotic handling

station (Thermo). Asynchronous cells and S-phase accumulated cells, using a 24 h treatment with 2 mM HU, were analysed. On-bead digest and mass spectrometry were performed as described[60]. Data represents three independent experiments for each condition, analysed by back-to-back MS and quantified by Label free quantification (LFQ). Proteins were identified and quantified with the MaxQuant 1.5 software suite by searching against the Uniprot human database. M(ox) and protein N-terminal acetylation were set as variable, and carbamylation as a fixed modification, with a 1 % FDR. Contaminants and reverse data base hits were deleted. Protein significantly enriched by GFP-DONSON were selected on the basis of p-value <0.05, and >2 fold change from asynchronous to S-phase, as identified by Student t-test and ratio cut-off against the respective negative control LFQ data as determined by MaxQuant ($p<0.05$; 2-fold).

**Proximity ligation assay (PLA).**

PLA was carried out as described in[5,56]. Briefly, cells from GFP or GFP-DONSON Flp-In T-REx HeLa cell lines were treated with 1ug/ml doxycycline and fixed/extracted after 24 h. For PCNA visualisation, cells were fixed with methanol at -20 °C for 10 min followed by a 5 min extraction in 0.3% Triton-X100 in PBS. For RPA visualisation, cells were pre-extracted in nuclear extraction buffer for 5 minutes on ice, and fixed in 3.6% paraformaldehyde for 10 minutes at room temperature. Cells were then incubated in anti-PCNA (PC10, 1:500, Santa Cruz) or anti-RPA (NA18; 1:500; Merck-Millipore) antibodies along with anti-GFP antibody (ab6556, 1:500, Abcam), and *in situ* proximity ligation was performed using a Duolink Detection Kit (Sigma Aldrich). Nuclear foci were imaged using a Nikon Eclipse Ni-U microscope equipped with a 100X oil lens in conjunction with a Zyla camera, and images were acquired using Elements software (Nikon). More than 200 cells were analysed per experiment per condition.

**Fluorescence Cross-Correlation Spectroscopy (FCCS)**

HeLa cells stably expressing GFP-DONSON and mCherry-PCNA (construct kindly provided by C. Lukas, Copenhagen; referred to as RFP-PCNA) were used for FCCS. For all details on Fluorescence Microscopy Imaging and FCS/FCCS, refer to **Supplementary Note**.

**Statistical Analyses**

Statistical differences were analyzed by: two-tailed Student T-Test (**Fig. 3b, 3d, 3f, 4f, 4g, 5b-d, 5f, 6a, 6e, 6f, 7b, 7d** and **Supplementary Fig. 5c, 8b, 9a, 12c, 13i, 14c, 16a-c, 17a-c, 19, 20, 21**); Mann-Whitney rank sum test (**Fig. 3e, 6c, 7c** and **Supplementary Fig. 12b, 15a, 15c**); and Chi-Squared Test (**Fig 4c, 6g** and **Supplementary Fig. 16d**). n refers to number of independent experiments unless indicated. Error bars represent standard error of the mean (s.e.m.) unless specified.

**Data Availability**

The NGS data used in the manuscript can be obtained from the European Genome-phenome Archive (EGA) under accession EGAS00001002224. NGS datasets on patients P14-P18, p21 are not available due to institutional IRB restrictions. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD005690.

**Methods-only references**


56. Higgs, M.R. *et al.* BOD1L Is Required to Suppress Deleterious Resection of Stressed Replication Forks. *Mol Cell* **59**, 462-77 (2015).

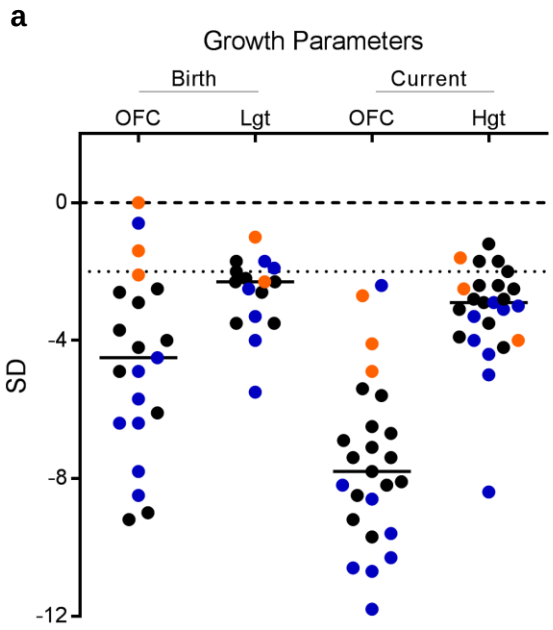
57. Singh, G. & Cooper, T.A. Minigene reporter for identification and analysis of cis elements and trans factors affecting pre-mRNA splicing. *Biotechniques* **41**, 177-81 (2006).

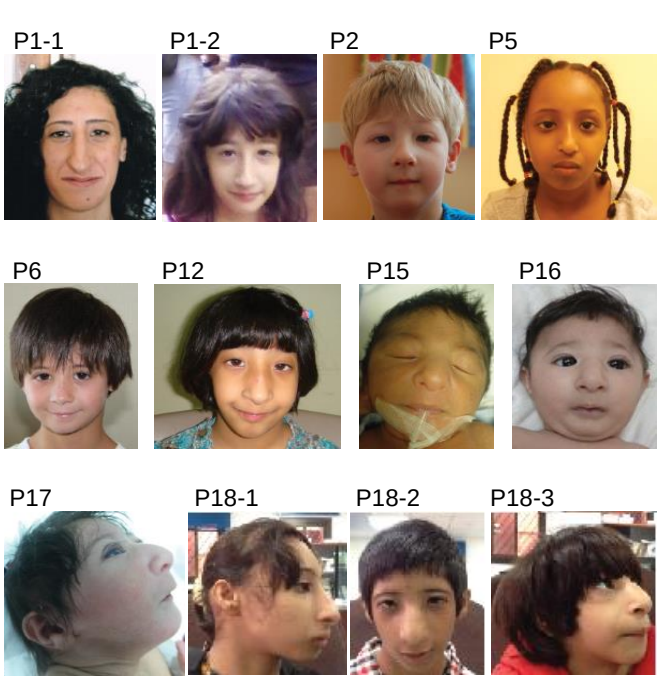
58. Sirbu, B.M., Couch, F.B. & Cortez, D. Monitoring the spatiotemporal dynamics of proteins at replication forks and in assembled chromatin using isolation of proteins on nascent DNA. *Nat Protoc* **7**, 594-605 (2012).

59. Conti, C. *et al.* Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. *Mol Biol Cell* **18**, 3059-67 (2007).

60. Turriziani, B. *et al.* On-beads digestion in conjunction with data-dependent mass spectrometry: a shortcut to quantitative and dynamic interaction proteomics. *Biology (Basel)* **3**, 320-32 (2014).

**Figure 1**


a
Growth Parameters
Birth
Current
OFC
Lgt
OFC
Hgt
0
-4
-8
-12
SD


b
P1-1
P1-2
P2
P5
P6
P12
P15
P16
P17
P18-1
P18-2
P18-3

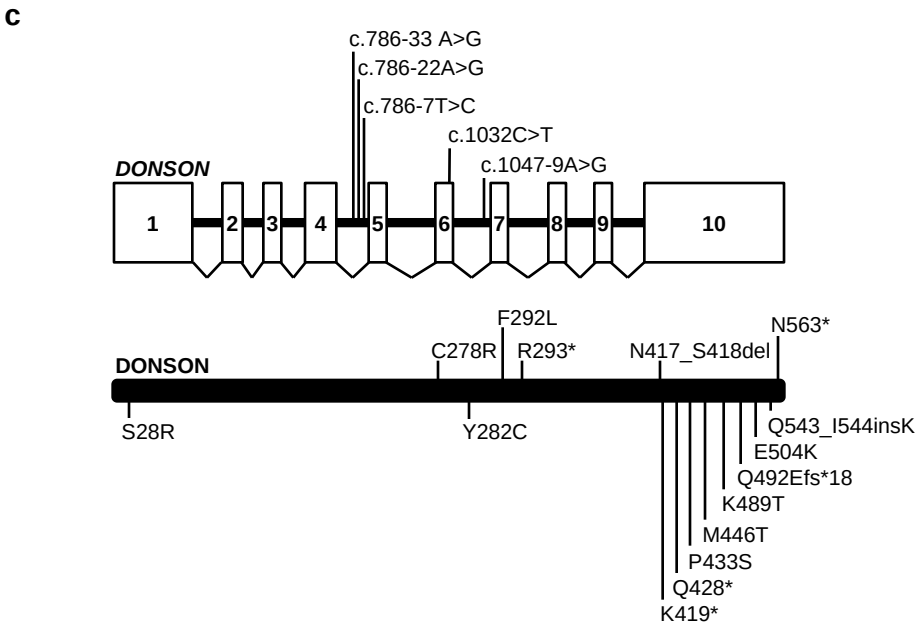

c
c.786-33 A>G
c.786-22A>G
c.786-7T>C
c.1032C>T
c.1047-9A>G
DONSON
1
2
3
4
5
6
7
8
9
10
DONSON
F292L
C278R
R293*
N417_S418del
N563*
S28R
Y282C
Q543_I544insK
E504K
Q492Efs*18
K489T
M446T
P433S
Q428*
K419*

**Figure 2**

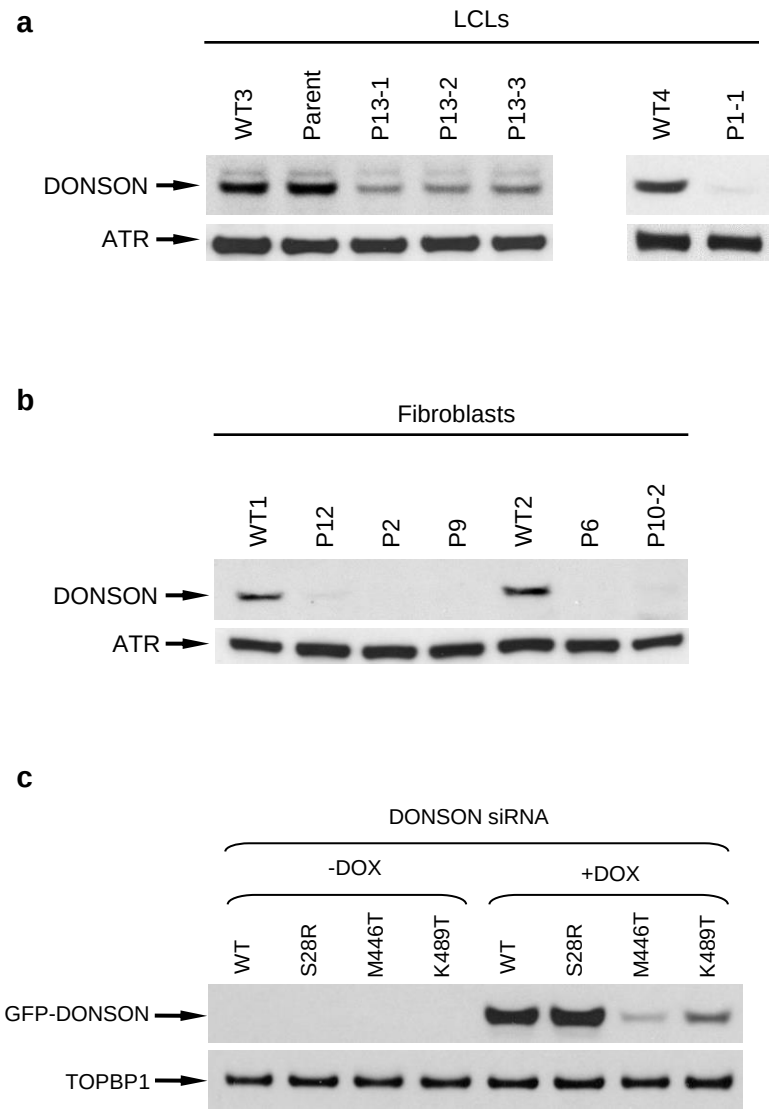

a
LCLs
WT3
Parent
P13-1
P13-2
P13-3
WT4
P1-1
DONSON
ATR
b
Fibroblasts
WT1
P12
P2
P9
WT2
P6
P10-2
DONSON
ATR
c
DONSON siRNA
-DOX
+DOX
WT
S28R
M446T
K489T
WT
S28R
M446T
K489T
GFP-DONSON
TOPBP1

**Figure 3**

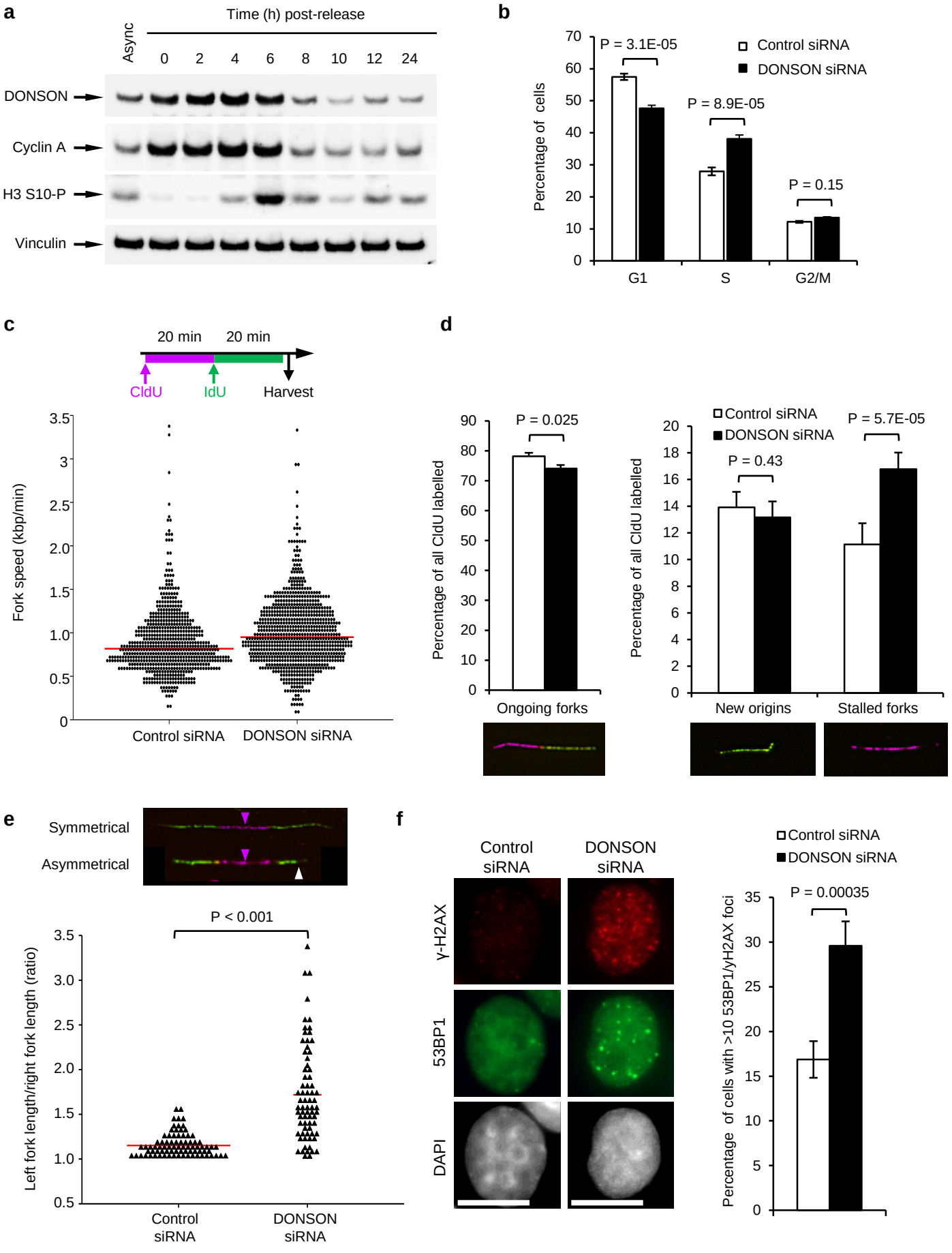

a
Time (h) post-release
Async
0
2
4
6
8
10
12
24
DONSON
Cyclin A
H3 S10-P
Vinculin
b
Percentage of cells
P = 3.1E-05
P = 8.9E-05
P = 0.15
Control siRNA
DONSON siRNA
G1
S
G2/M
c
20 min
20 min
CldU
IdU
Harvest
Fork speed (kbp/min)
Control siRNA
DONSON siRNA
d
Percentage of all CldU labelled
P = 0.025
Ongoing forks
Control siRNA
DONSON siRNA
P = 0.43
P = 5.7E-05
New origins
Stalled forks
e
Symmetrical
Asymmetrical
P < 0.001
Left fork length/right fork length (ratio)
Control siRNA
DONSON siRNA
f
Control siRNA
DONSON siRNA
γ-H2AX
53BP1
DAPI
Control siRNA
DONSON siRNA
P = 0.00035
Percentage of cells with >10 53BP1/γH2AX foci

Figure 4

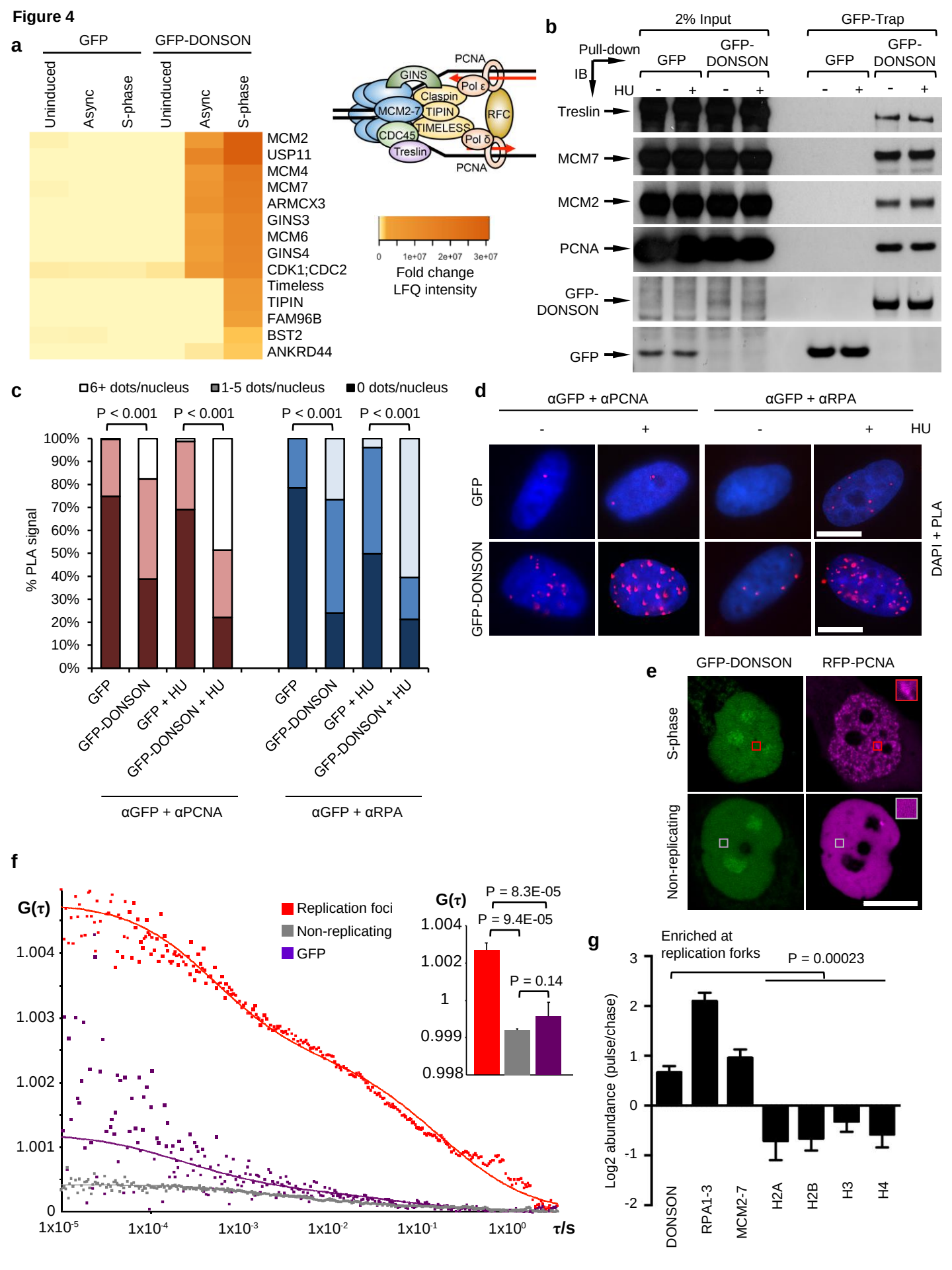

a
GFP
GFP-DONSON
Uninduced
Async
S-phase
Uninduced
Async
S-phase
MCM2
USP11
MCM4
MCM7
ARMCX3
GINS3
MCM6
GINS4
CDK1;CDC2
Timeless
TIPIN
FAM96B
BST2
ANKRD44
PCNA
GINS
Pol ε
Claspin
MCM2-7
TIPIN
RFC
TIMELESS
CDC45
Pol δ
Treslin
PCNA
0
1e+07
2e+07
3e+07
Fold change
LFQ intensity
b
2% Input
GFP-Trap
Pull-down
IB
GFP
GFP-DONSON
GFP
GFP-DONSON
HU
- + - + - + - +
Treslin
MCM7
MCM2
PCNA
GFP-DONSON
GFP
c
6+ dots/nucleus
1-5 dots/nucleus
0 dots/nucleus
P < 0.001
P < 0.001
P < 0.001
P < 0.001
% PLA signal
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
GFP
GFP-DONSON
GFP + HU
GFP-DONSON + HU
GFP
GFP-DONSON
GFP + HU
GFP-DONSON + HU
αGFP + αPCNA
αGFP + αRPA
d
αGFP + αPCNA
αGFP + αRPA
- + - + HU
GFP
GFP-DONSON
DAPI + PLA
e
GFP-DONSON
RFP-PCNA
S-phase
Non-replicating
f
G(τ)
Replication foci
Non-replicating
GFP
1.004
1.003
1.002
1.001
0
1x10-5
1x10-4
1x10-3
1x10-2
1x10-1
1x100
τ/s
G(τ)
P = 8.3E-05
P = 9.4E-05
P = 0.14
1.004
1.002
1
0.999
0.998
g
Enriched at
replication forks
P = 0.00023
Log2 abundance (pulse/chase)
3
2
1
0
-1
-2
DONSON
RPA1-3
MCM2-7
H2A
H2B
H3
H4

Figure 5

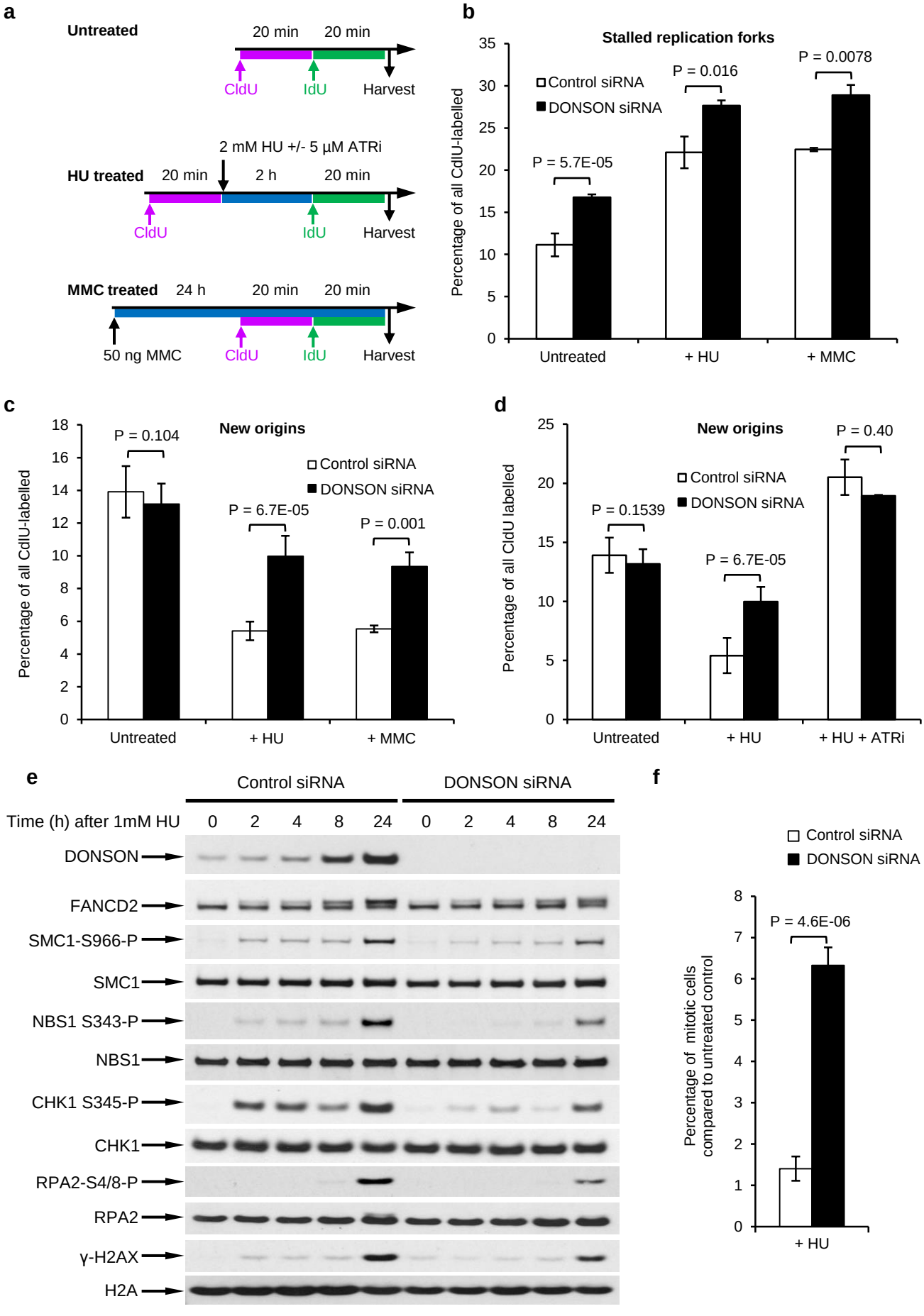

a
Untreated
20 min
20 min
CldU
IdU
Harvest
2 mM HU +/- 5 μM ATRi
HU treated
20 min
2 h
20 min
CldU
IdU
Harvest
MMC treated
24 h
20 min
20 min
50 ng MMC
CldU
IdU
Harvest
b
Stalled replication forks
Control siRNA
DONSON siRNA
P = 5.7E-05
P = 0.016
P = 0.0078
Percentage of all CdIU-labelled
Untreated
+ HU
+ MMC
c
New origins
Control siRNA
DONSON siRNA
P = 0.104
P = 6.7E-05
P = 0.001
Percentage of all CdIU-labelled
Untreated
+ HU
+ MMC
d
New origins
Control siRNA
DONSON siRNA
P = 0.1539
P = 6.7E-05
P = 0.40
Percentage of all CldU labelled
Untreated
+ HU
+ HU + ATRi
e
Control siRNA
DONSON siRNA
Time (h) after 1mM HU
0 2 4 8 24 0 2 4 8 24
DONSON
FANCD2
SMC1-S966-P
SMC1
NBS1 S343-P
NBS1
CHK1 S345-P
CHK1
RPA2-S4/8-P
RPA2
γ-H2AX
H2A
f
Control siRNA
DONSON siRNA
P = 4.6E-06
Percentage of mitotic cells compared to untreated control
+ HU

**Figure 6**

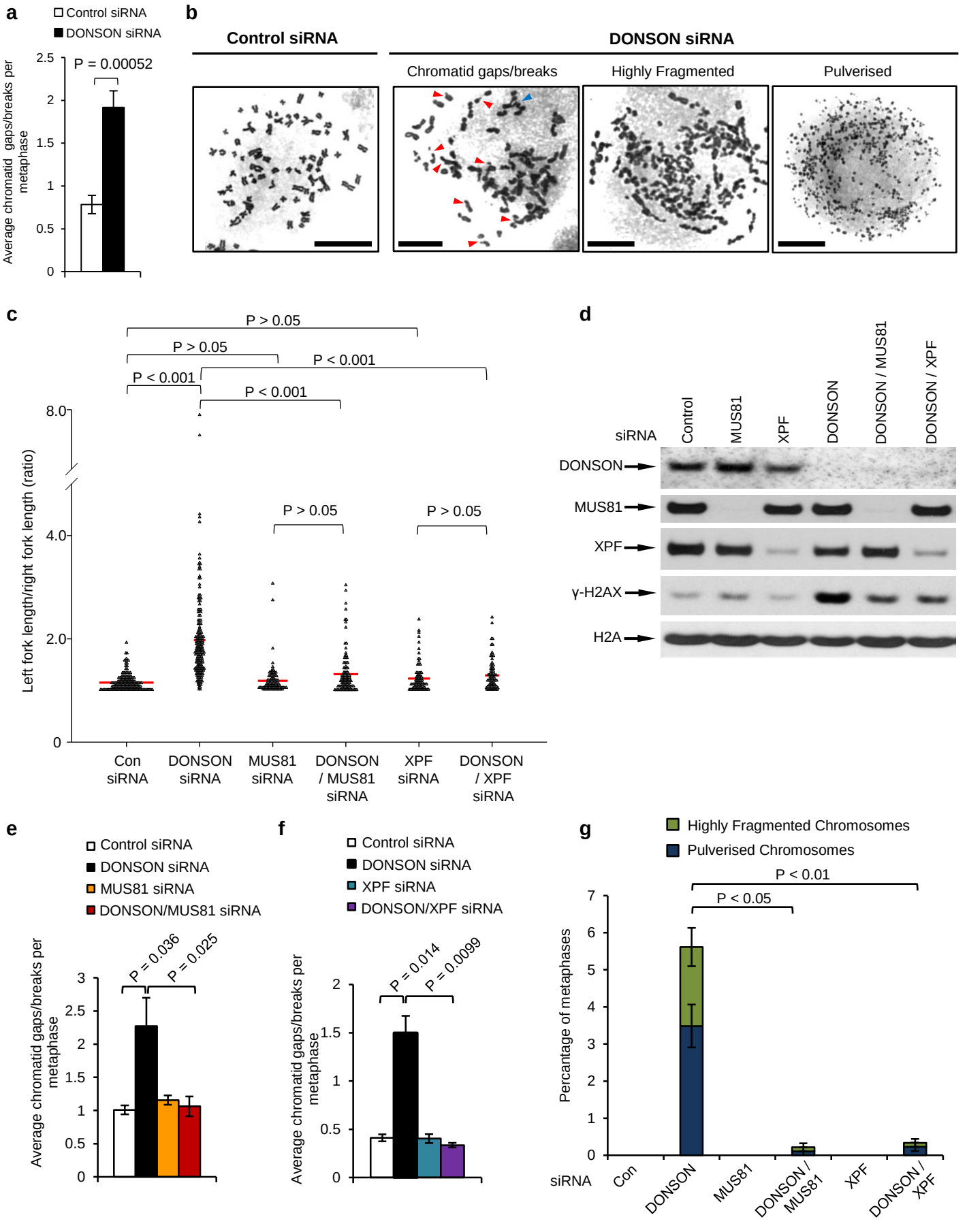
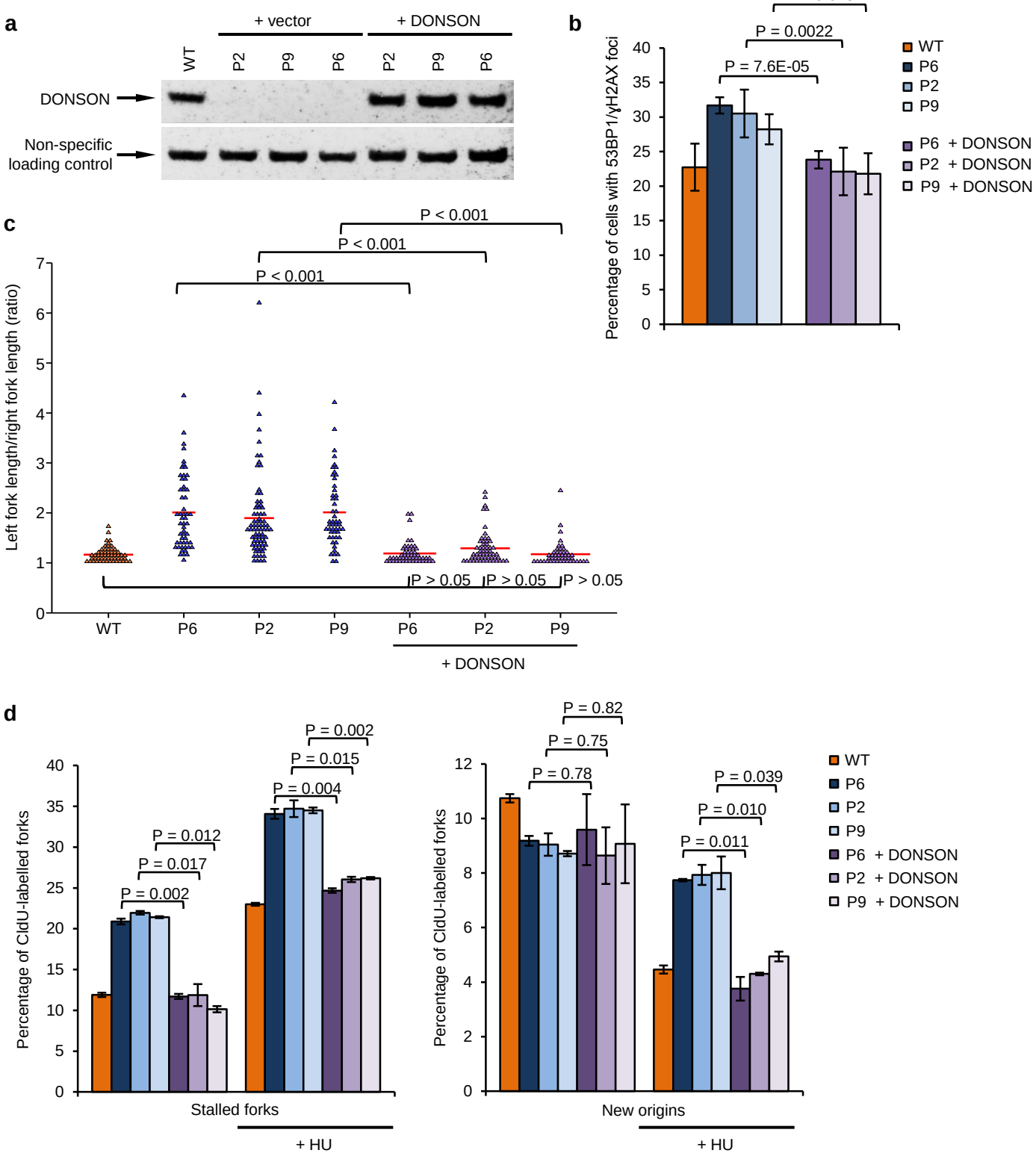

a
Control siRNA
DONSON siRNA
P = 0.00052
Average chromatid gaps/breaks per metaphase
b
Control siRNA
DONSON siRNA
Chromatid gaps/breaks
Highly Fragmented
Pulverised
c
P > 0.05
P > 0.05
P < 0.001
P < 0.001
P < 0.001
P > 0.05
P > 0.05
Left fork length/right fork length (ratio)
Con siRNA
DONSON siRNA
MUS81 siRNA
DONSON / MUS81 siRNA
XPF siRNA
DONSON / XPF siRNA
d
siRNA
Control
MUS81
XPF
DONSON
DONSON / MUS81
DONSON / XPF
DONSON
MUS81
XPF
γ-H2AX
H2A
e
Control siRNA
DONSON siRNA
MUS81 siRNA
DONSON/MUS81 siRNA
P = 0.036
P = 0.025
Average chromatid gaps/breaks per metaphase
f
Control siRNA
DONSON siRNA
XPF siRNA
DONSON/XPF siRNA
P = 0.014
P = 0.0099
Average chromatid gaps/breaks per metaphase
g
Highly Fragmented Chromosomes
Pulverised Chromosomes
P < 0.01
P < 0.05
Percantage of metaphases
siRNA
Con
DONSON
MUS81
DONSON / MUS81
XPF
DONSON / XPF

**Figure 7**


a
+ vector
+ DONSON
WT
P2
P9
P6
P2
P9
P6
DONSON
Non-specific loading control
b
P = 0.025
P = 0.0022
P = 7.6E-05
Percentage of cells with 53BP1/γH2AX foci
WT
P6
P2
P9
P6 + DONSON
P2 + DONSON
P9 + DONSON
c
P < 0.001
P < 0.001
P < 0.001
Left fork length/right fork length (ratio)
P > 0.05
P > 0.05
P > 0.05
WT
P6
P2
P9
P6
P2
P9
+ DONSON
d
P = 0.002
P = 0.012
P = 0.017
P = 0.002
P = 0.015
P = 0.004
Percentage of CldU-labelled forks
Stalled forks
+ HU
P = 0.82
P = 0.75
P = 0.78
P = 0.039
P = 0.010
P = 0.011
Percentage of CldU-labelled forks
New origins
+ HU
WT
P6
P2
P9
P6 + DONSON
P2 + DONSON
P9 + DONSON

**Supplementary Information**

# Mutations in *DONSON* disrupt replication fork stability and cause microcephalic dwarfism.

**Supplementary Table 1: Clinical phenotype data of individuals with DONSON mutations**

Clinical information relating to individuals with *DONSON* mutations. Wgt, weight (kg); Hgt, height (cm); Lgth, length (cm); SD, standard deviation from the population mean for age and sex; OFC, occipital frontal circumference; GORD, gastro-oesophageal reflux disease; ID, intellectual disability; ADHD, attention deficit hyperactivity disorder; FA, Fanconi anaemia. [a]measurement taken at 5 weeks of age. [b]measurement taken at 4 months of age. [c]measurement taken at 2 weeks of age. [d]term gestation assumed for SD calculation. [e]fetal death. [f]antenatal USS scan. [g]at 19 wk scan, standard deviation calculated using Fenton growth chart[61].

**Supplementary References:**

61. Fenton, T.R. & Kim, J.H. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatr* **13**, 59 (2013).

**Supplementary Table 2. Common ancestral haplotype (blue) flanking the DONSON locus in families P1, P2, P3 and P6.**

| SNP | position[b] | gene | coding | Ref | Alt | P1-1 | P1-2 | P2 | P3 | P6 | MAF[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs121913215 | 34804930 | *IFNGR2* | intronic | A | T | AT | AT | AT | AT | AA | 0.0010 |
| rs75398504 | 34839530 | *TMEM50B* | intronic | G | A | GA | GA | GA | GA | GG | 0.0006 |
| rs144266803 | 34861444 | *DNAJC28* | T86I | G | A | GA | GA | GA | GA | GA | 0.0010 |
| rs148155071 | 34876289 | *GART* | 3'UTR | A | T | AT | AT | AT | AT | AT | 0.0026 |
| rs151206034 | 34889888 | *GART* | M577T | A | G | AG | AG | AG | AG | AG | 0.0010 |
| rs139633075 | 34897281 | *GART* | L365V | G | C | GC | GC | GC | GC | GC | 0.0010 |
| rs187476997 | 34902228 | *GART* | intronic | T | C | TC | TC | TC | TC | TC | 0.0016 |
| rs142038738 | 34903861 | *GART* | E177D | C | A | CA | CA | CA | CA | CA | 0.0009 |
| rs148794591 | 34924549 | *SON* | A1004A | T | C | TC | TC | TC | TC | TC | 0.0010 |
| rs142751481 | 34925936 | *SON* | I1467V | A | G | AG | AG | AG | AG | AG | 0.0010 |
| rs146664036 | 34951753 | ***DONSON*** | **K489T** | T | G | TG | TG | TG | TG | TG | 0.0010 |
| rs192585552 | 34956005 | ***DONSON*** | **c.786-33** | T | C | TC | TC | TC | TC | TC | 0.0005 |
| no rsID[a] | 34960866 | ***DONSON*** | **S28R** | T | G | TG | TG | TG | TG | TG | - |
| rs190141427 | 34967690 | *CRYZL1* | 3'UTR | T | G | TG | TG | TG | TG | TG | 0.0006 |
| rs184229251 | 34989145 | *CRYZL1* | intronic | T | C | TC | TC | TC | TC | TC | 0.0006 |
| rs570210988 | 36261887 | *RUNX1* | intronic | C | C | CC | CC | CT | CC | CC | 0.0002 |

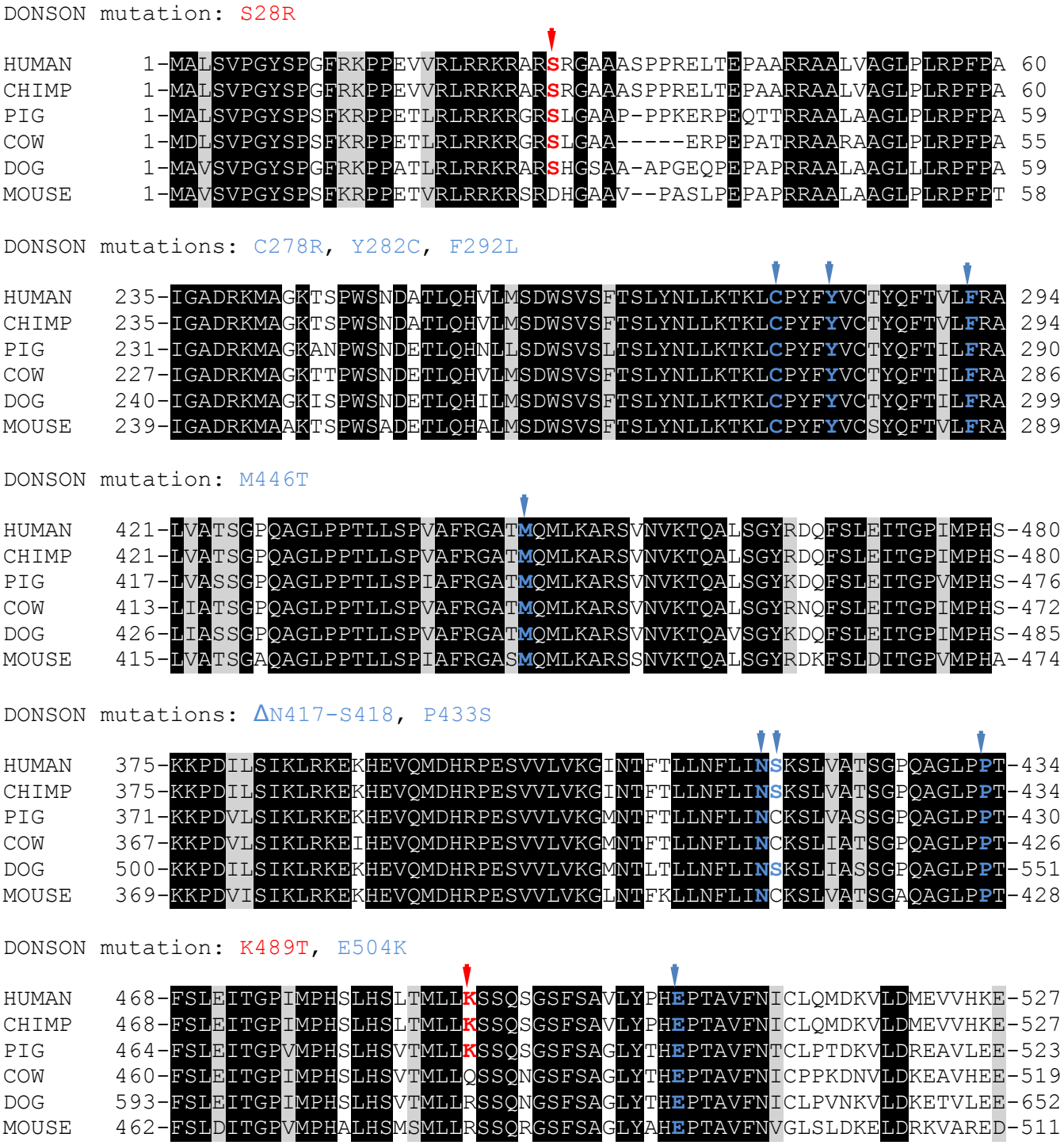
Analysis of rare variants identified through exome sequencing demonstrated a common ancestral haplotype extending 127.7 kb in DONSON patients heterozygous for the 'haplotype' disease allele. Parents of P2 and P3 were also exome sequenced, allowing phasing of alleles across this haplotype. No other variants were identified in trans in the genes within this region in any individuals apart from DONSON – the in trans variants in DONSON are omitted here for clarity. [a]This variant lies within a GC rich region and therefore is not covered in many control population datasets. [b]Genomic position using hg19 coordinates. [c]Minor allele frequencies taken from ExAC or 1000 Genomes data.

**Supplementary Table 3 – Proteomic mass spectrometry screen for GFP-DONSON interactors.**

DONSON proteomics screen (related to **Fig. 4a**). Cell extracts of GFP or GFP-DONSON HeLa Flp-In/T-Rex cells were prepared (-Tet; uninduced cells: Asynch; asynchronously growing cells: S-phase; cells treated with 2 mM HU to arrest cells in S-phase). GFP or GFP-DONSON was isolated and digested on-beads. Peptides were identified by mass spectrometry and quantified by LFQ by the MaxQuant software package with a 1% FDR. Contaminants and reverse data base hits were deleted. Sheet 1: GFP-DONSON protein interactors. Proteins enriched by GFP-DONSON were identified by Student t-test and ratio cut-off against the respective negative control ($p<0.05$; 2-fold) LFQ data as determined by MaxQuant. Sheet 2: GFP-DONSON-binding proteins enriched in the S-phase cells ($p<0.02$ & ratio>2. Q-value: quality score. Peptides: number of peptides identified. MS/MS count: number of independent spectra identified.

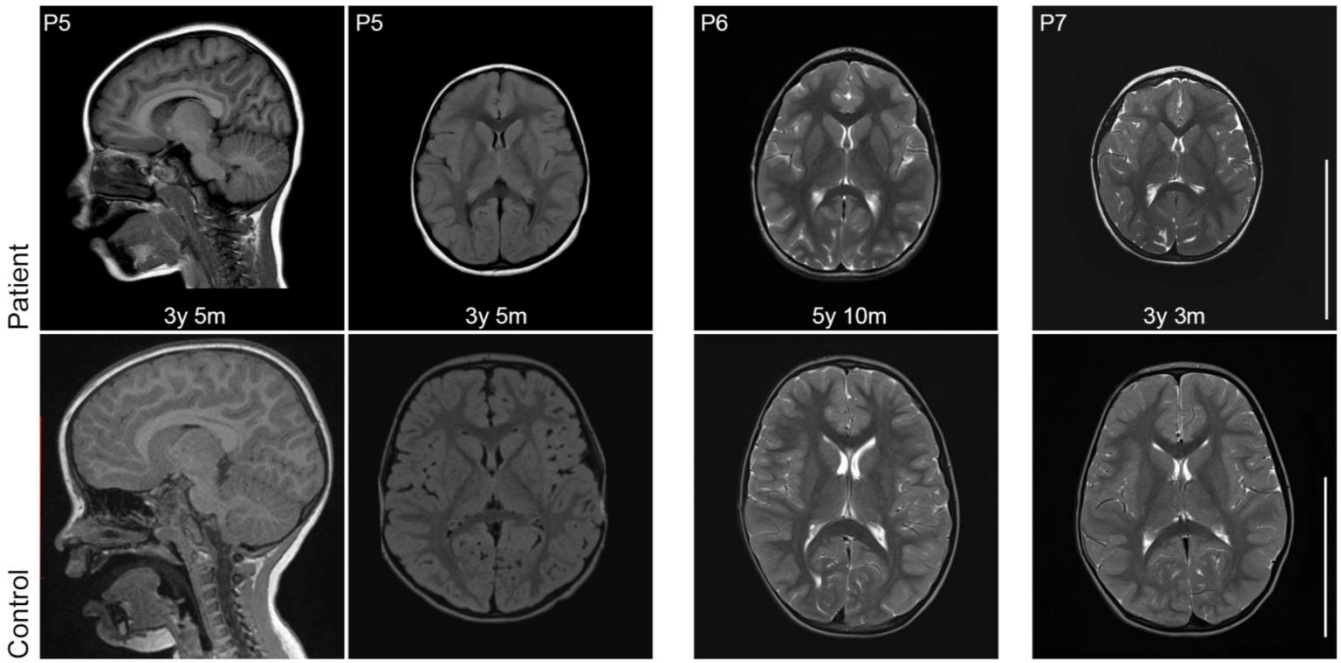
**Supplementary Table 4 – DONSON primer sequences used in this study**

Details of the DONSON primers used in this study



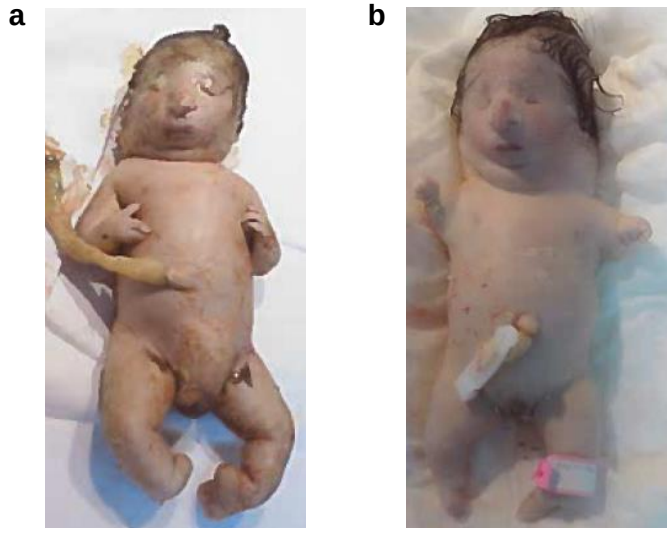

**Supplementary Figure 1: Conservation of DONSON amino acids mutated in MD patients.**

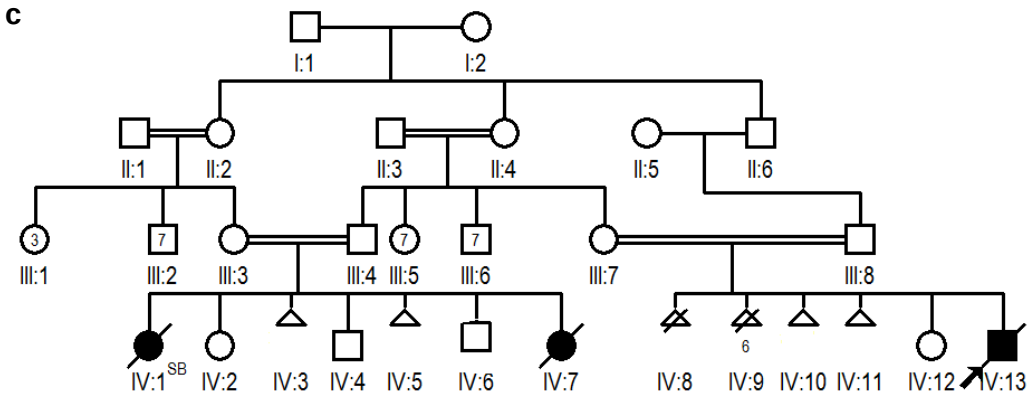
Amino acid alignment of DONSON protein from different species showing the degree of evolutionary conservation of disease causing DONSON point mutations, generated using Clustal Omega. Red arrows indicate the amino acid residues of DONSON mutated in MD patients. The K489T and S28R missense variants associated with the haplotype present in patients P1-7 are coloured red. All other mutations are coloured blue.

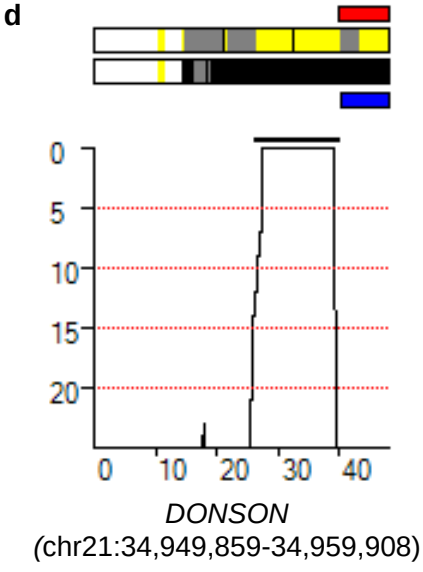


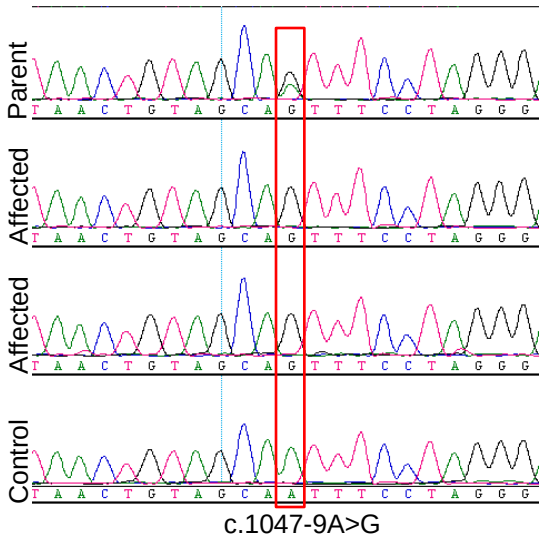

**Supplementary Figure 2: Cerebral cortical size is markedly reduced with simplification of gyral folding in patients with *DONSON* mutations.**

Magnetic resonance imaging T1-weighted sagittal and axial FLAIR images of patient P5, and T2-weighted sagittal and axial images of patients P6 and P7, compared with age-matched control scans of healthy individuals. Scale bars, 10 cm.



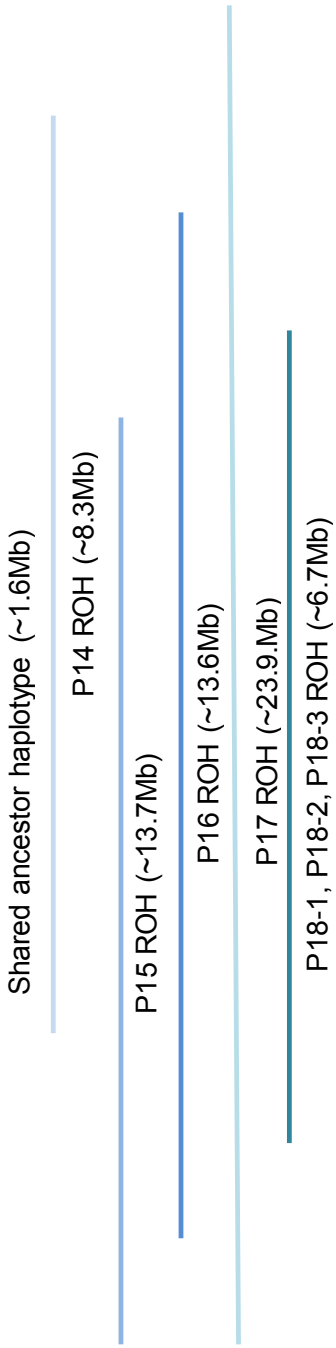




**Supplementary Figure 3: Clinical and genetic characterisation of affected individuals from family P21.**

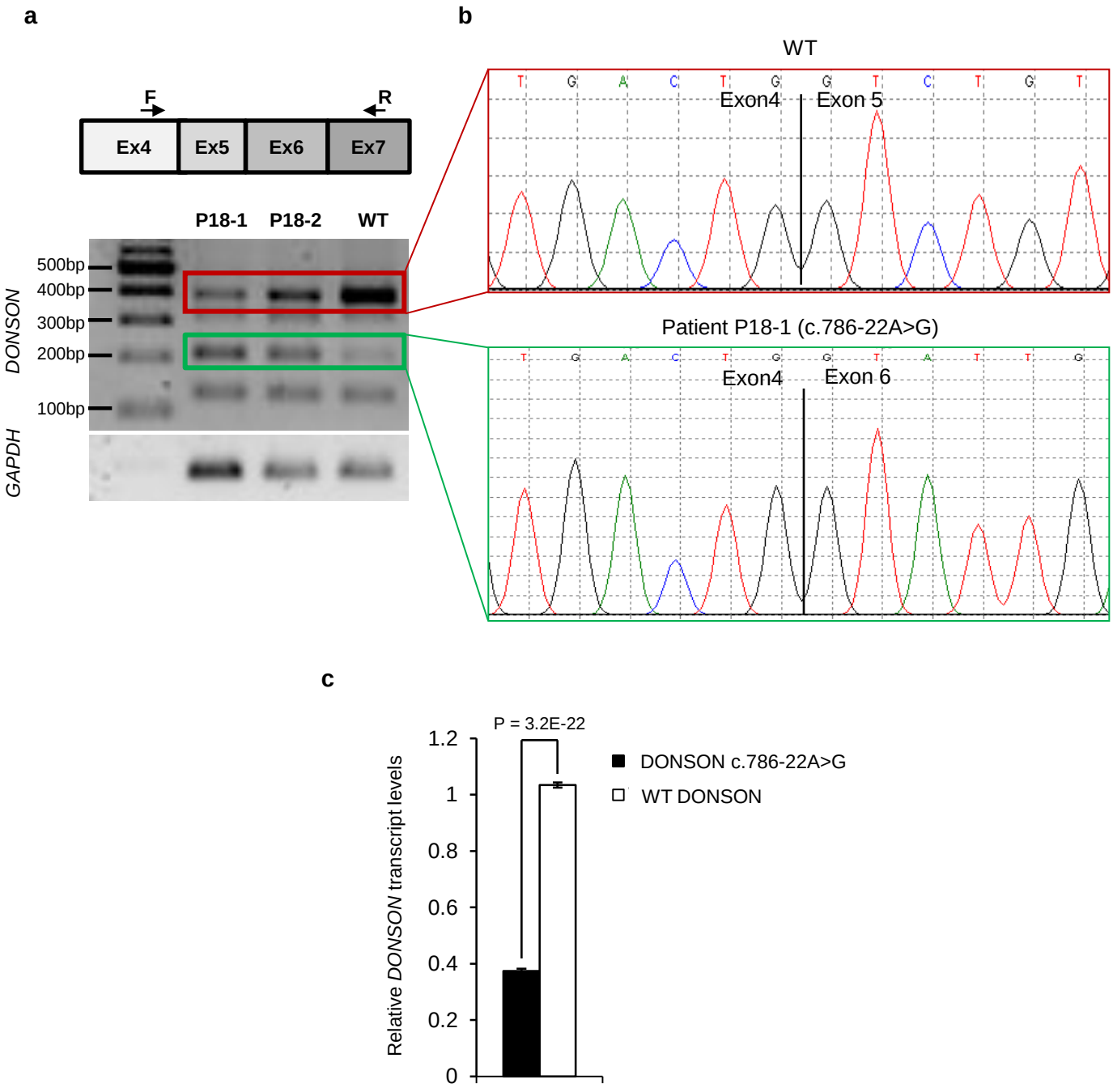
**(a)** and **(b)** are clinical photographs of dead fetuses (P21-1 and P21-2). **(c)** Pedigree of family P21. **(d)** IBDelphi image of chromosome 21. **(e)** Chromatograms indicating the presence of the c.1047-9A>G mutation within both affected fetuses, and a parent, from family P21.

| SNPs | Distance Mb | P14 | P15 | P16 | P17 | P18-1 | P18-2 | P18-3 |
|---|---|---|---|---|---|---|---|---|
| RS7279441 | 24.1Mb | AA | AA | AA | ab | AA | AA | AA |
| RS1118467 | 24.2Mb | ab | ab | BB | AA | ab | ab | ab |
| 1332 markers | | | | | | | | |
| RS171487 | 29Mb | ab | BB | BB | BB | ab | ab | ab |
| RS2831134 | 29Mb | AA | AA | AA | AA | ab | ab | ab |
| 142 markers | | | | | | | | |
| RS2831813 | 29.8Mb | BB | BB | ab | BB | BB | BB | BB |
| RS4420778 | 29.8Mb | BB | BB | BB | BB | BB | BB | BB |
| 519 markers | | | | | | | | |
| RS2255507 | 32.4Mb | BB | AA | AA | BB | ab | ab | ab |
| RS2833282 | 32.4Mb | AA | AA | AA | AA | AA | AA | AA |
| 440 markers | | | | | | | | |
| RS17632819 | 34.4Mb | BB | ab | BB | AA | BB | BB | BB |
| RS2409469 | 34.4Mb | BB | BB | BB | BB | BB | BB | BB |
| 67 markers | | | | | | | | |
| RS17860260 | 34.6Mb | BB | BB | BB | BB | AA | AA | AA |
| RS2834167 | 34.6Mb | BB | BB | BB | BB | BB | BB | BB |
| RS2515716 | 34.6Mb | AA | AA | AA | AA | AA | AA | AA |
| 33 markers | | | | | | | | |
| RS2834236 | 34.9Mb | AA | AA | AA | AA | AA | AA | AA |
| *DONSON* | 34.9Mb | AA | AA | AA | AA | AA | AA | AA |
| DONSON | 34.9Mb | BB | BB | BB | BB | BB | BB | BB |
| RS2070392 | 35Mb | BB | BB | BB | BB | BB | BB | BB |
| 280 markers | | | | | | | | |
| RS2834657 | 36.2Mb | AA | AA | AA | AA | AA | AA | AA |
| RS2834658 | 36.2Mb | BB | BB | BB | BB | BB | BB | BB |
| RS2298352 | 36.2Mb | BB | BB | AA | BB | BB | BB | BB |
| 280 markers | | | | | | | | |
| RS2835192 | 37.3Mb | AA | AA | AA | AA | AA | AA | AA |
| RS11910121 | 37.3Mb | ab | BB | BB | BB | BB | BB | BB |
| 326 markers | | | | | | | | |
| RS2211845 | 39.1Mb | AA | AA | AA | BB | BB | BB | BB |
| RS8133218 | 39.1Mb | BB | BB | BB | BB | ab | ab | ab |
| 1253 markers | | | | | | | | |
| RS150792 | 43.4Mb | BB | BB | BB | BB | BB | BB | BB |
| RS150796 | 43.4Mb | ab | AA | ab | BB | BB | BB | BB |
| 436 markers | | | | | | | | |
| RS4819274 | 48Mb | BB | BB | BB | AA | BB | ab | BB |
| RS9637 231 | 48Mb | AA | AA | ab | BB | AA | ab | AA |



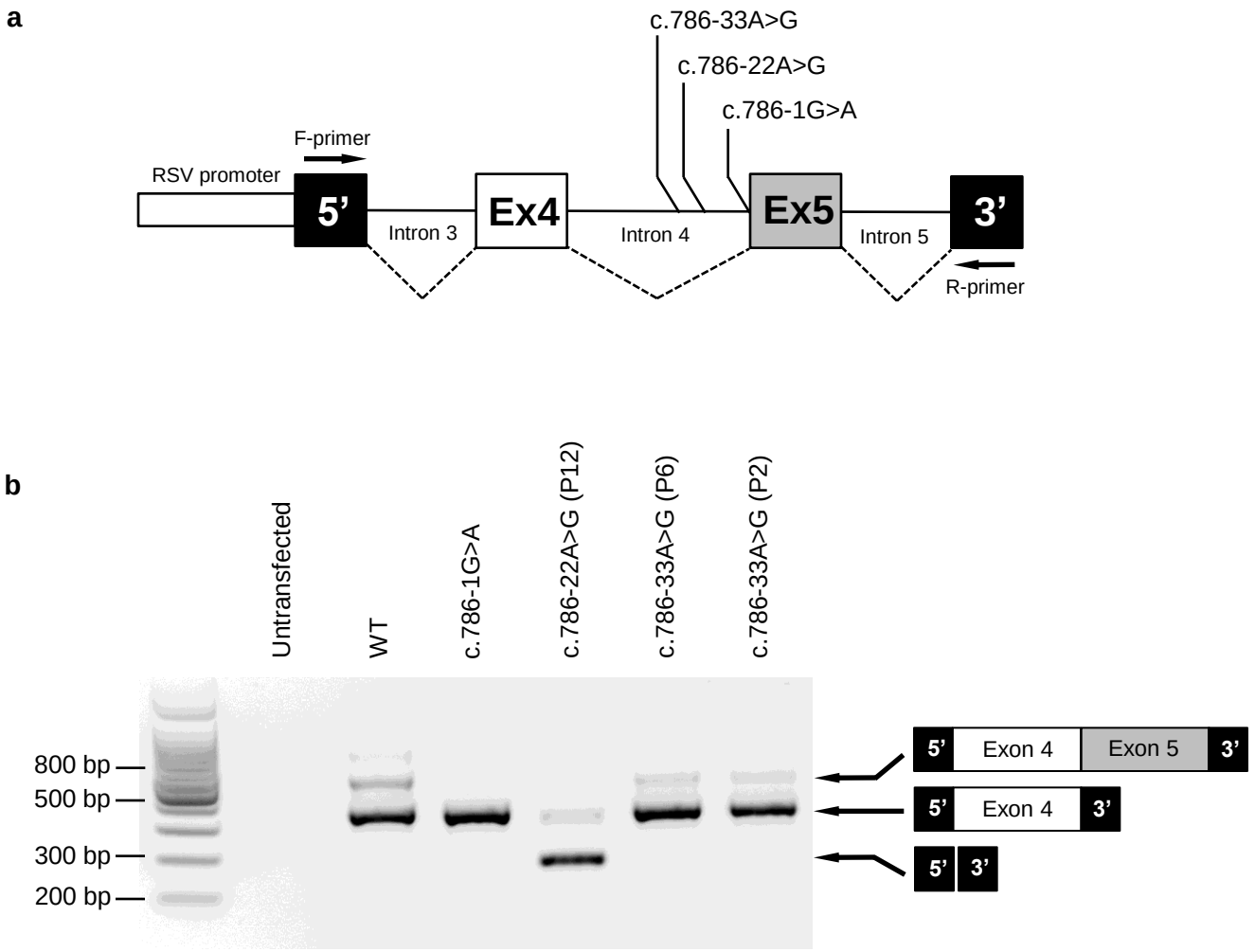

**Supplementary Figure 4: Common ancestral haplotype flanking the *DONSON* locus in families P14 to P18-3.**

Patients P14, P15, P16, P17, P18-1, P18-2 and P18-3 have a distant common ancestry, sharing a common ~1.6 Mb homozygous haplotype across the *DONSON* locus. Schematic of high density SNP genotyping of the *DONSON* locus in the six patients indicated above. The different families are indicated by different shades of blue. Heterozygous SNP markers delineating regions of homozygosity are shown in red. An identical haplotype of 321 SNPs is evident within the 1.6 Mb region of homozygosity in the six patients, which is consistent with a shared common ancestor (indicated by grey shading). 313 homozygous SNP markers within the 1.6 Mb regions are omitted for clarity.



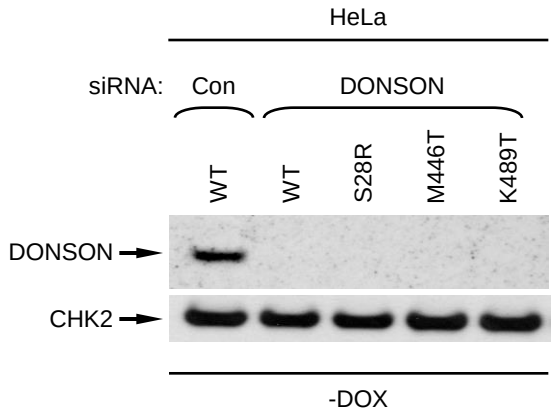

**Supplementary Figure 5: The patient-associated *DONSON* intronic mutation c.786-22A>G affects mRNA splicing.**

**(a)** Top: a schematic of *DONSON* mRNA with the position of the PCR primers (F = forward primer; R = reverse primer) indicated relative to the exons. Bottom: PCR amplification of *DONSON* exons 4-7 from cDNA derived from a WT individual or patients 18-1 and 18-2. PCR amplification of *GAPDH* was used as a control. **(b)** Sequencing electropherogram of the PCR products highlighted in (a). The red box indicates the WT sequence. The green box indicates skipping of exon 5 of *DONSON*. **(c)** Quantification of the relative amounts of the WT *DONSON* mRNA as compared to the levels of GAPDH by RT-qPCR from three normal individuals and four *DONSON* patients (n=3).



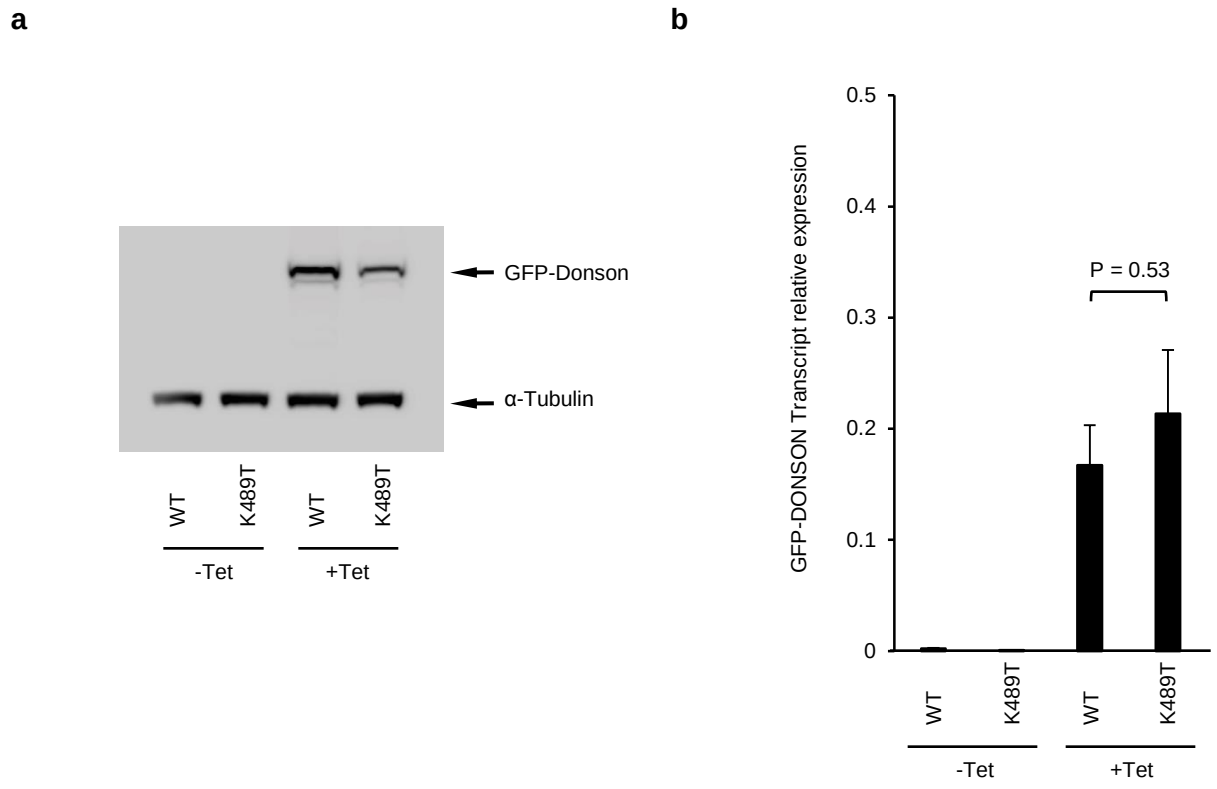

**Supplementary Figure 6: The patient-associated *DONSON* intronic mutation c.786-33A>G does not affect mRNA splicing.**

**(a)** A schematic of *DONSON* mRNA splicing mini gene reporter construct with the relative positions of the PCR primers, RSV promoter, two artificial vector-associated exons (labelled as 5' and 3') and *DONSON* exons/introns indicated. The location of the two *DONSON* intronic variants (c.786-22A>G and c.786-33A>G) relative to the intron/exon boundaries are shown. An essential splice site c.786-1G>A mutation was generated as a positive control that disrupts splicing. The position of the PCR primers (F-primer = forward primer; R-primer = reverse primer) is indicated relative to the exons. **(b)** PCR amplification of *DONSON* exons 4-5 from cDNA derived from the splicing reporter construct containing WT or the individual *DONSON* mutations. The exon content of each PCR product was verified by sequencing (indicated).



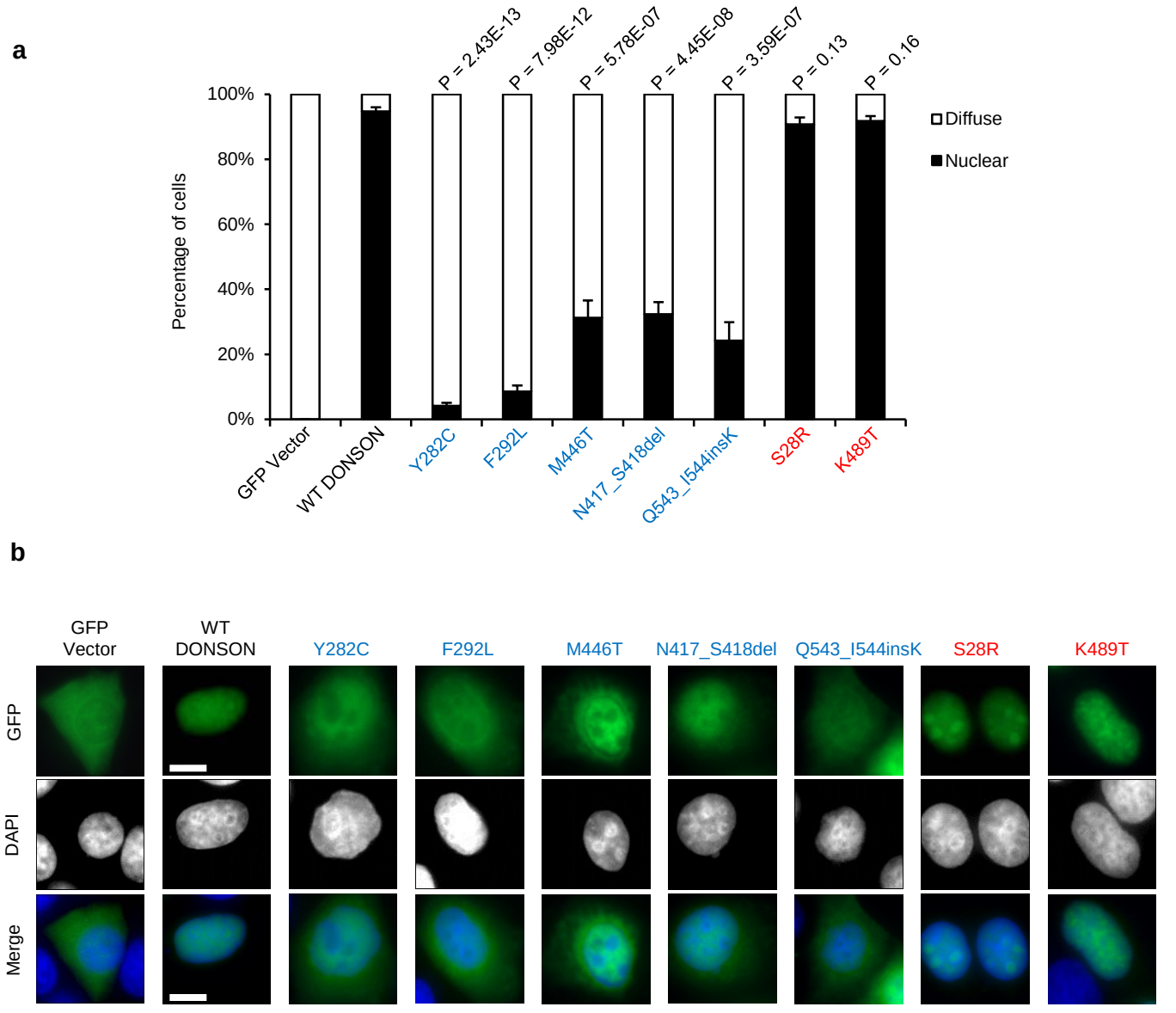

**Supplementary Figure 7: Depletion of endogenous DONSON in HeLa Flp-In/T-Rex cells**

Endogenous DONSON was depleted using siRNA from HeLa Flp-In/T-Rex cells expressing doxycycline-inducible GFP-tagged siRNA-resistant WT or mutant DONSON. Endogenous DONSON were detected using an anti-DONSON antibody. CHK2 was used as a loading control.



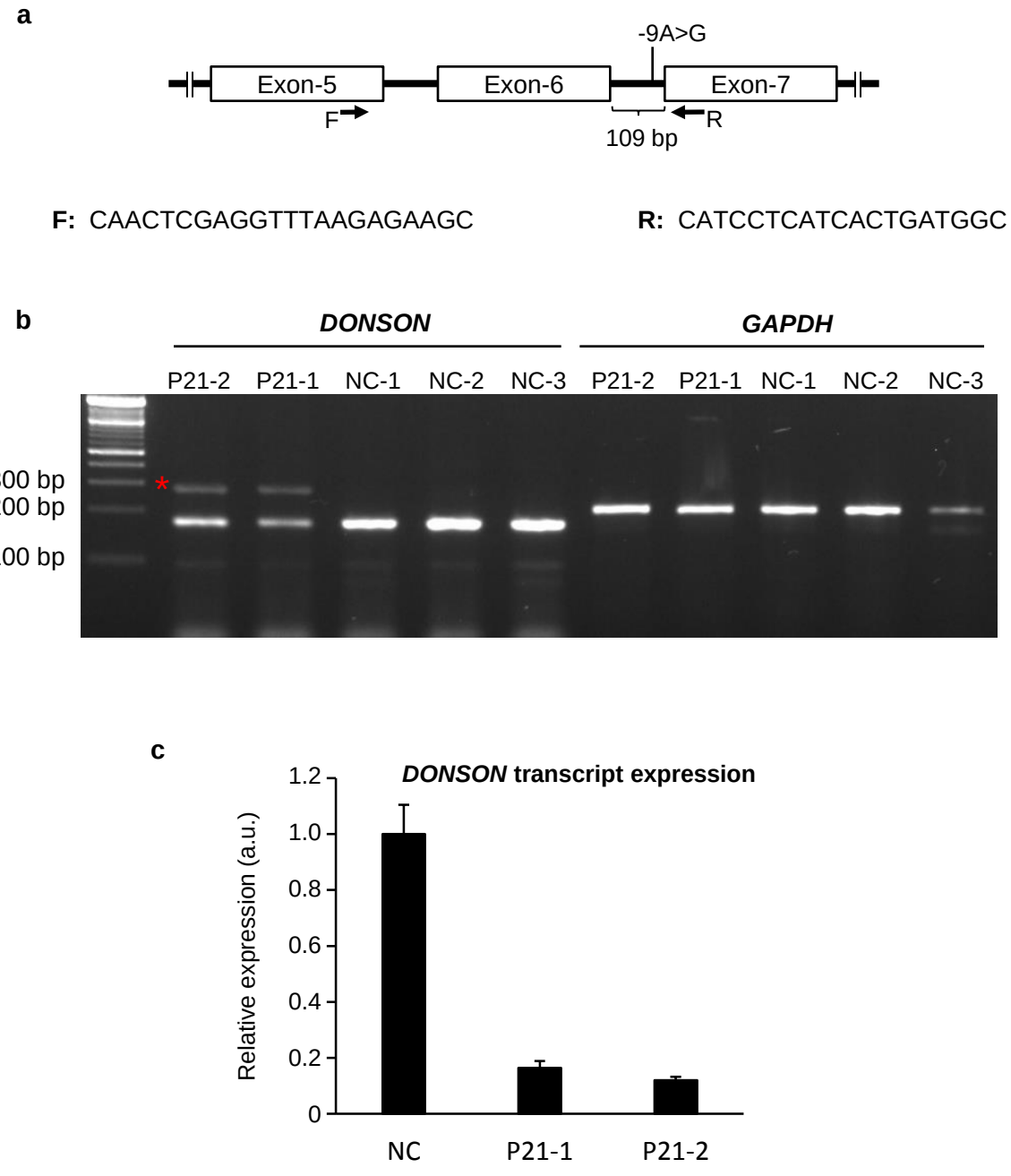

**Supplementary Figure 8: Alterations in the protein level of the K489T GFP-DONSON variant are not due to reduced levels of mRNA**

**(a)** Immunoblotting using an anti-GFP antibody demonstrates reduced protein levels of the K489T GFP-DONSON variant compared to WT GFP-DONSON (same cell lines as **Fig 2d**). '+Tet', expression induced by treatment with 1 ug/ml tetracycline for 24 h; '-Tet' uninduced. α-Tubulin was used as a loading control. **(b)** qRT-PCR analysis of GFP-DONSON transcript levels from RNA isolated from the same cell pellets as in (a). Transcript levels are expressed relative to the housekeeping gene PBGD. Data in (a) and (b) are representative of three technical replicates from two independent experiments.



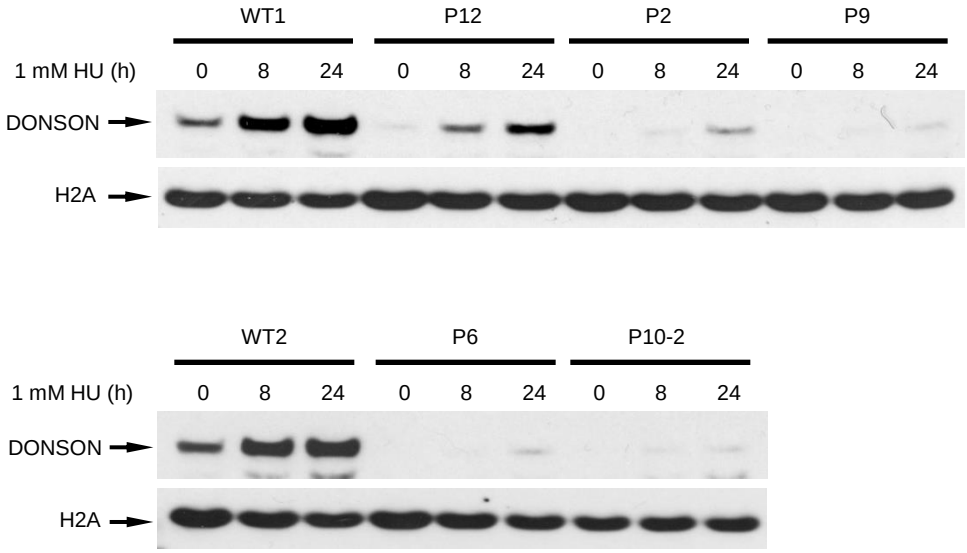

**Supplementary Figure 9: The patient-associated DONSON point mutations disrupt protein sub-cellular localisation.**

**(a)** Patient-associated point mutations in DONSON affect its subcellular localization. HeLa cells were transfected with a vector encoding either WT GFP-DONSON, or GFP-DONSON containing patient-derived mutations and fixed 24 h post-transfection. The percentages of GFP positive cells that contained either a solely nuclear GFP signal, or a diffuse pan-cellular GFP signal were quantified (n=5). Statistical differences between WT and mutated GFP-DONSON were determined by Student's t-test. **(b)** Representative images are shown. Missense mutations associated with the haplotype in patients P1 to P7 are shaded red and all other mutations are shaded blue. Scale bars; 10 µm.



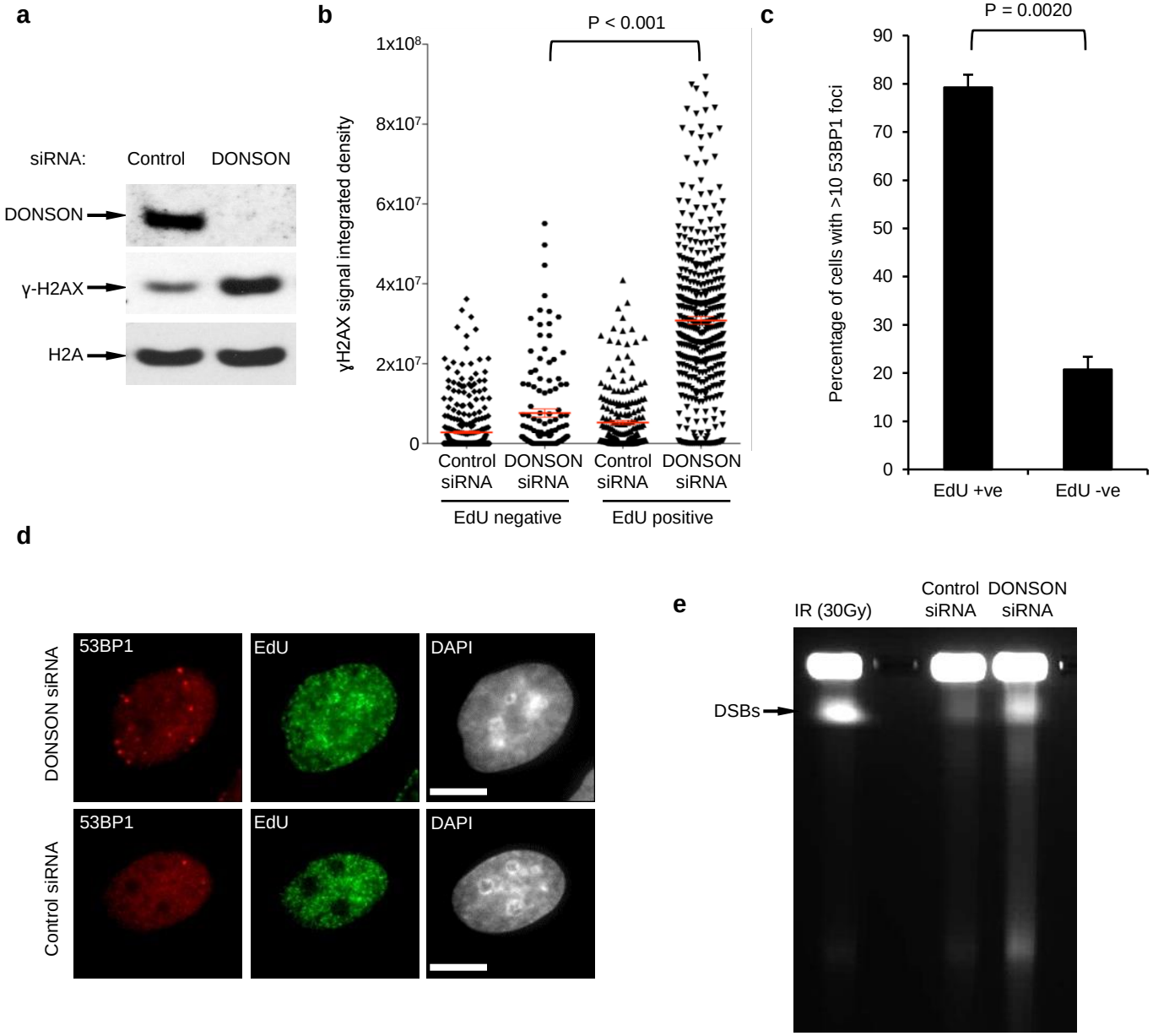

**Supplementary Figure 10: The patient-associated *DONSON* intronic mutation c.1047-9A>G affects mRNA abundance.**

**(a)** Schematic presentation of exon and intron boundaries of a region of *DONSON*. Arrows indicates primers (F; forward: and R; reverse) used for semi-quantitative RT-PCR and qRT-PCR. **(b)** Semi-quantitative RT-PCR analysis of *DONSON* transcript levels in P21-1 and P21-2 and three normal controls (NC-1, NC-2 and NC-3). Red star indicates retention of intronic region in patient amplicon. **(c)** qRT-PCR analysis of *DONSON* transcript levels, normalised to *GAPDH*, in patients P21-1, P21-2 and three normal controls (NC) (n=3). Error bars indicate standard deviation (a.u. = arbitary units).



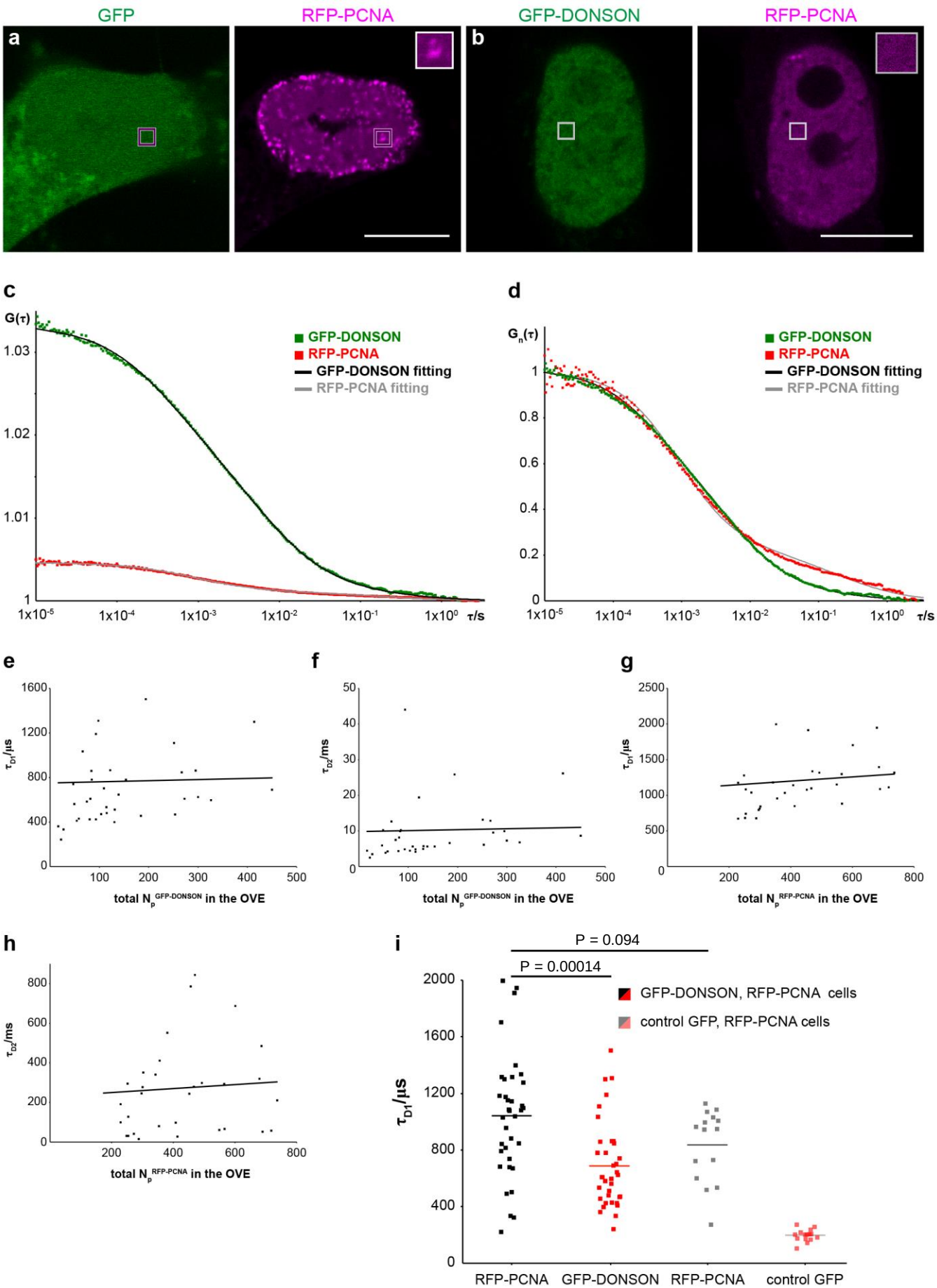

**Supplementary Figure 11: Patient derived fibroblasts retain residual levels of DONSON protein.**

Immunoblotting of cell extracts from patient derived fibroblast cell lines treated with HU for the indicated times. H2A was used as a loading control.




**Supplementary Figure 12: Depletion of DONSON gives rise to increased spontaneous DNA damage in replicating cells.**

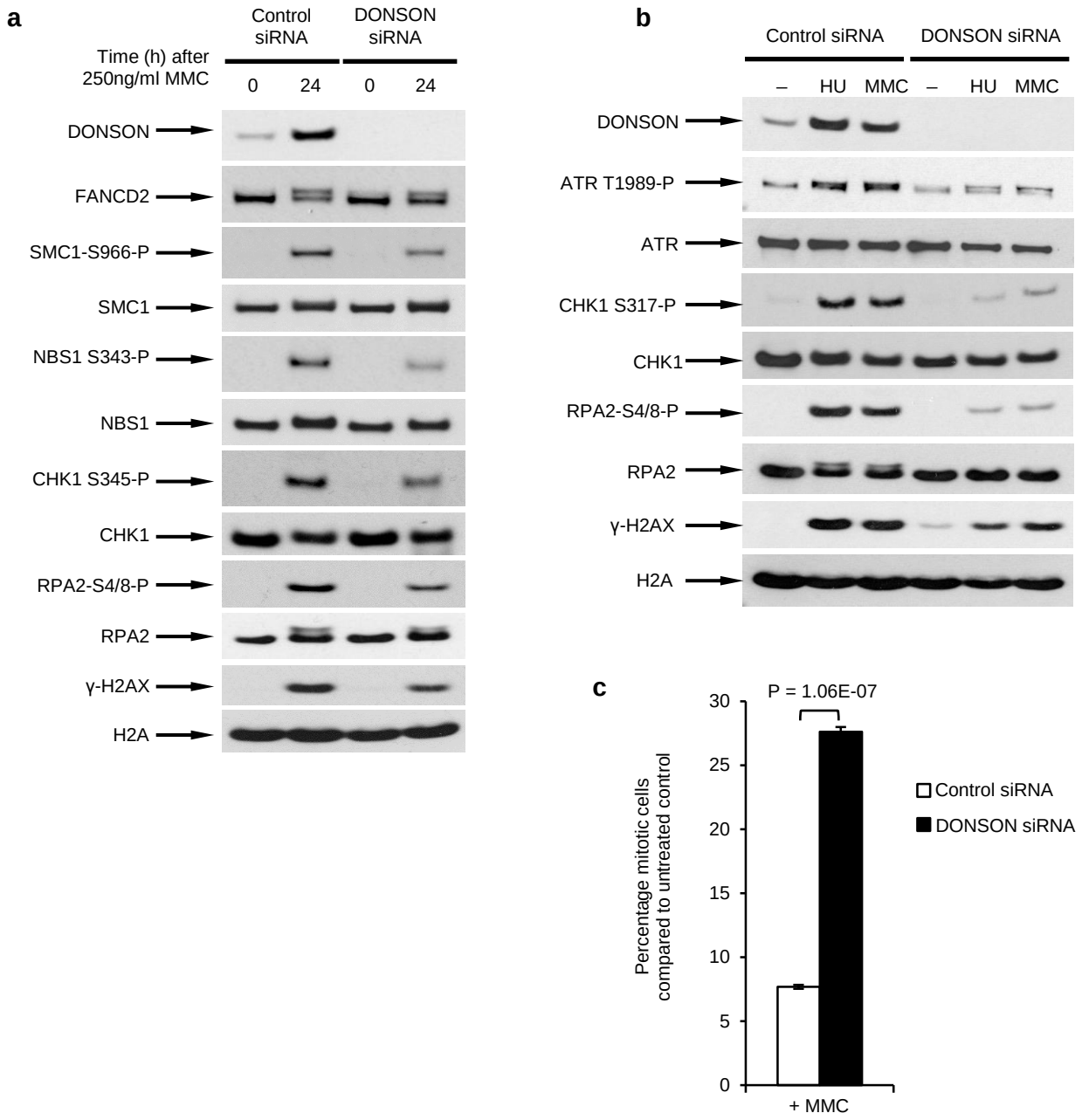
**(a)** Knockdown of DONSON results in increased levels of phosphorylated H2AX. HeLa cells were transfected with control or DONSON siRNA, harvested 72 h post-transfection and extracts subjected to SDS-PAGE and immunoblotting using the antibodies indicated (n=3). **(b-e)** DONSON depletion causes DNA double strand breaks in replicating cells. **(b)** DONSON-depleted cells were seeded onto coverslips and incubated with 10 μM EdU for 10 min before harvesting. Cells were then pre-extracted, fixed and stained with an antibody to γH2AX. EdU was visualised using the Click-iT-Alexa-488 Imaging Kit. The intensity of γH2AX immunostaining in >100 EdU positive cells and >50 EdU negative cells were quantified per cell line per condition per experiment (n=2). **(c)** Quantification of the percentage of EdU positive and negative cells with >10 53BP1 foci in cells without DONSON. A minimum of 300 cells were counted per sample per independent experiment. (n=3). **(d)** Representative images for (c) Scale bars; 10 μm. **(e)** Pulsed-field gel electrophoresis was used to detect the presence of DNA double strand breaks (DSBs) in cells from (a). As a positive control, HeLa cells were irradiated with 30 Gy of ionising radiation (IR) and harvested 1 h later.




**Supplementary Figure 13: Fluorescence Cross-Correlation Spectroscopy of GFP-DONSON and RFP-PCNA.** See next page for legend.

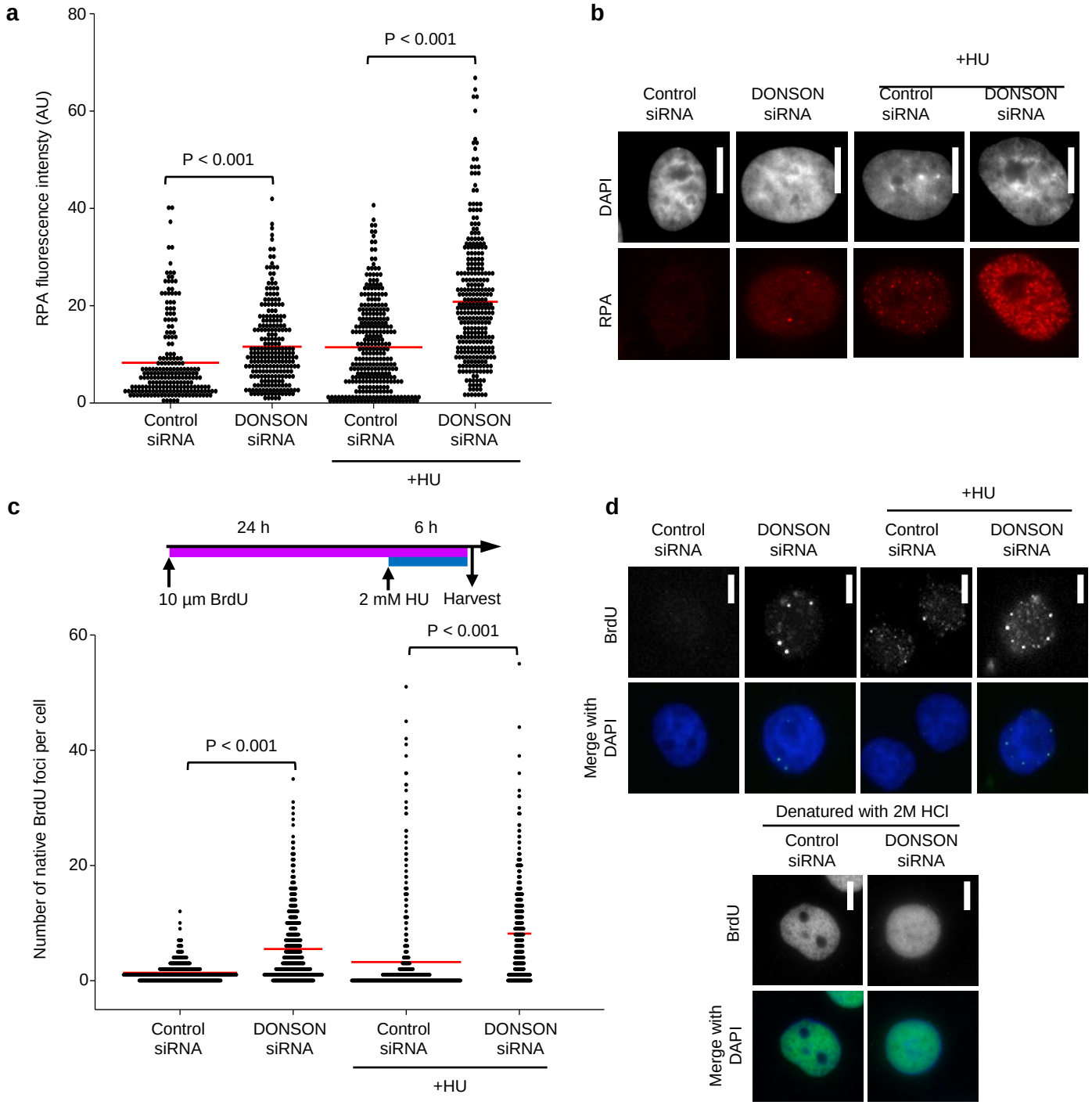
**Supplementary Figure 13: Fluorescence Cross-Correlation Spectroscopy of GFP-DONSON and RFP-PCNA.**

Fluorescence Cross-Correlation Spectroscopy (FCCS) is a method used to analyse the degree of interaction of two differently coloured fluorescent proteins randomly diffusing through a fixed volume. Control measurements for GFP, GFP-DONSON and RFP-PCNA from FCS(fluorescence Correlation Spectroscopy)/FCCS experiments. **(a, b)** Representative confocal images of HeLa cells expressing **(a)** GFP and RFP-PCNA in S-phase, and **(b)** GFP-DONSON and RFP-PCNA in non-S phase. The purple and grey boxes indicate regions representative of those in which FCS/FCCS measurement were performed. Scale bars, 10 μm. **(c)** Average Auto-Correlation Curves (ACCs) of measurements of GFP-DONSON and RFP-PCNA in replication foci (displayed in **Fig. 4e, f**). High concentrations of both RFP-PCNA and GFP-DONSON (3 ± 1.2 μM and 1.1 ± 0.8 μM, respectively, n=36) were measured by FCS at replication foci, even in cells displaying low average amounts of PCNA and DONSON. For cross-correlation analysis, (**Fig. 4f**) only measurements from cells displaying similar levels of RFP-PCNA, and GFP-DONSON or GFP, were compared. Fitting was performed using a two-component model with triplet formation (see **Supplementary Note 1**). **(d)** Average FCS measurements normalized to the same amplitude revealed a marked difference in the mobility of the bound fraction of each protein (represented by the slow component in the ACCs, with PCNA displaying slower movement, as would be expected from a protein encircling DNA), which can be observed by the shift of the RFP-PCNA ACC to longer characteristic times. We determined $\tau_{D_1}^{RFP-PCNA} = 1041 \pm 430\,\mu s$, $\tau_{D_2}^{RFP-PCNA} = 221 \pm 225\,ms$, $\tau_{D_1}^{GFP-DONSON} = 688 \pm 303\,ms$ and $\tau_{D_2}^{GFP-DONSON} = 24 \pm 34\,ms$ (n=36) in GFP-DONSON expressing cells. **(e-h)** Characteristic decay times $\tau_{D_1}$ and $\tau_{D_2}$ are not significantly altered upon concentration increase of total GFP-DONSON and RFP-PCNA molecules, as evident from $\tau_{D_1} = f\,(N_p)$ and $\tau_{D_2} = f\,(N_p)$, where $N_p$ is the average number of molecules of GFP-DONSON and RFP-PCNA and $\tau_{D_1}$, $\tau_{D_2}$ the diffusion times of the unbound and bound fractions, indicating that the observed fluorescence intensity fluctuations are generated by diffusion. **(i)** Comparison of the characteristic decay times $\tau_{D_1}$(fast FCS component, when fitted with a two-component model for three-dimensional diffusion and triplet formation) of RFP-PCNA and GFP-DONSON ($n$=36 measurements) or control GFP ($n$=15 measurements). RFP-PCNA showed considerably slower diffusion of the fast component than GFP-DONSON or control GFP, but no significant differences in its mobility were observed between cells expressing GFP-DONSON or control GFP (Student's unpaired two-tailed t-test between GFP-DONSON and RFP-PCNA, $p$=0.00014; and between RFP-PCNA across the two cell lines, $p$=0.094). Horizontal lines denote the average of measurements for each condition.



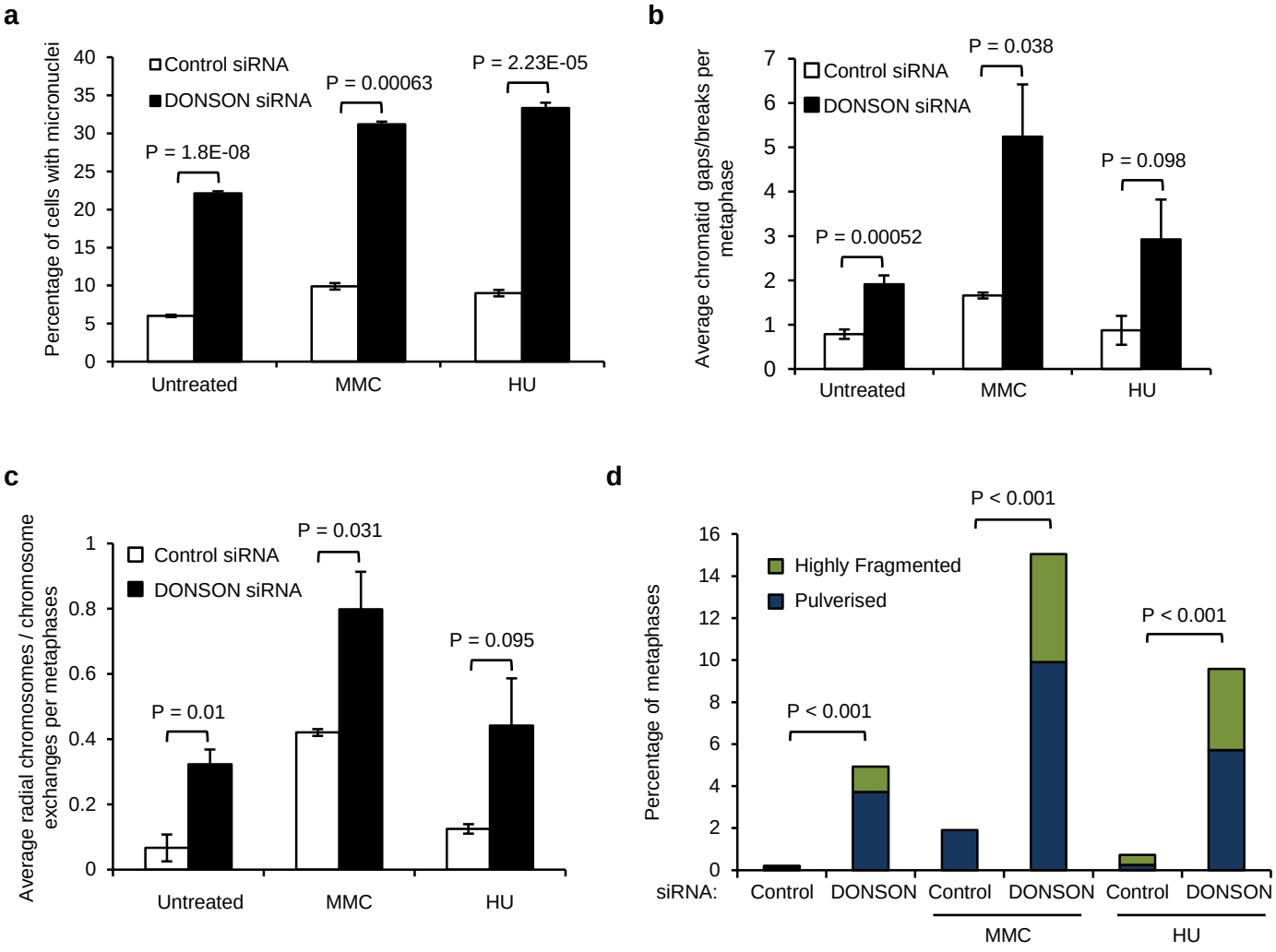

**Supplementary Figure 14: Reduced ATR-CHK1-dependent signalling and defective activation of the G2/M checkpoint in DONSON-depleted cells following replication stress treatment**

**(a)** Cells lacking DONSON exhibit defective or delayed ATR auto-phosphorylation in response to replication stress. HeLa cells transfected with either control or DONSON siRNA were exposed to 1 mM HU or 250 ng/ml MMC for 24 h, and subjected to immunoblotting using the indicated antibodies (n=2). **(b)** Cells lacking DONSON fail to efficiently activate the ATR-dependent replication stress response following exposure to MMC. Whole cell extracts of HeLa cells transfected with either control or DONSON siRNA were subjected to immunoblot analysis using the indicated antibodies following treatment with 250 ng/ml MMC for 24 h. **(c)** The percentage of mitotic cells in cells from (a) was determined by flow cytometry using antibodies to phosphorylated histone H3-Ser10 as a marker of mitotic cells. Data represents the mean of three independent experiments.



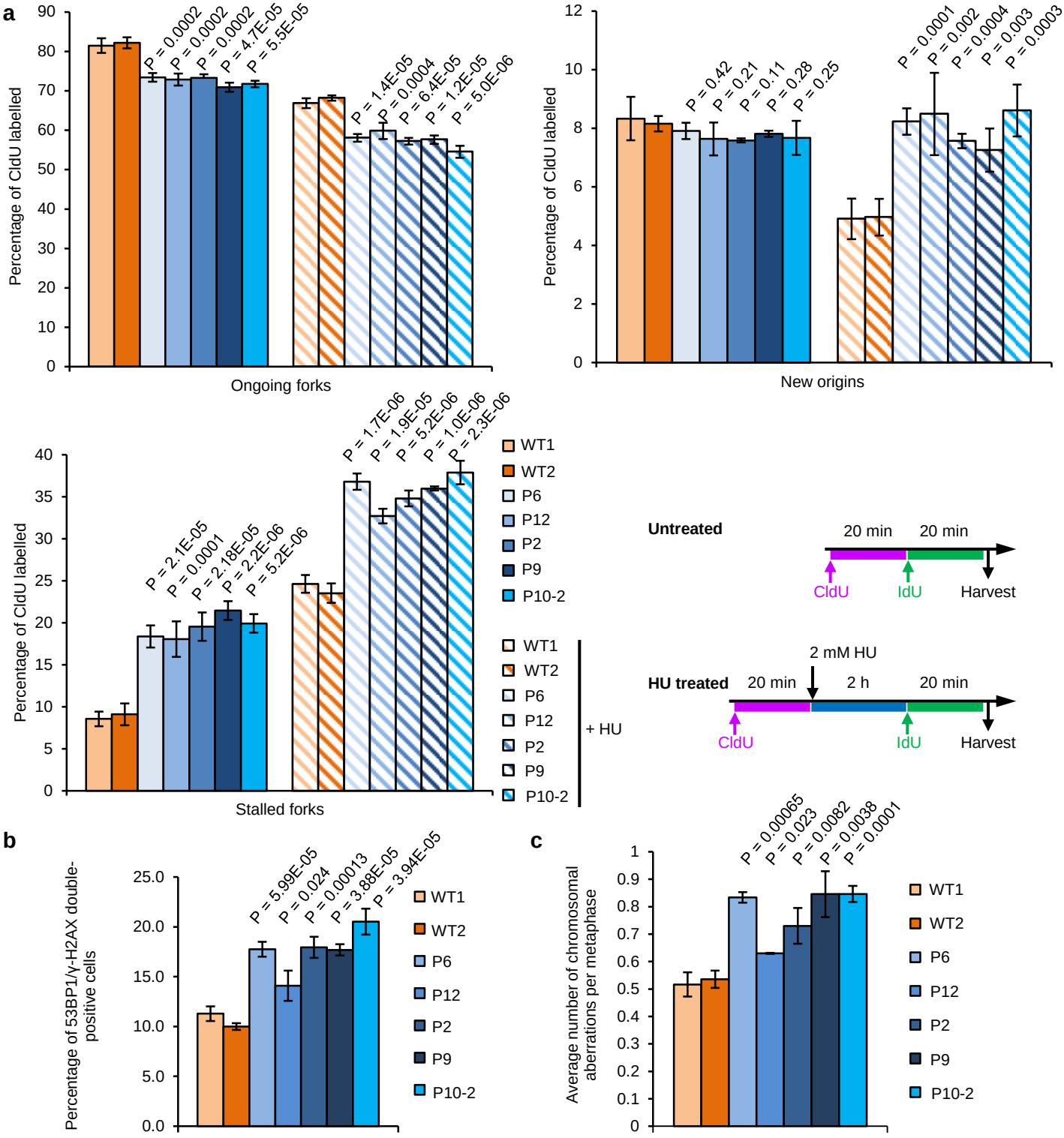

**Supplementary Figure 15: Increased levels of RPA-coated ssDNA in DONSON-depleted cells following treatment with HU**

Exposure of cells lacking DONSON to HU induces increased levels of RPA coated ssDNA. **(a-b)** HeLa cells transfected with either control or DONSON siRNA were exposed to 1 mM HU for 24 h, immunostained with antibodies specific to RPA2 and analysed by fluorescence microscopy. Fluorescence intensity per nucleus was quantified using ImageJ. Lines denote mean values from two independent experiments. Representative images are shown in **(b)**. **(c-d)** Native BrdU foci formation in HeLa cells transfected with the indicated siRNAs was analysed by fluorescence microscopy. The cells were treated as detailed in the schematic (top), and immunostained with antibodies to BrdU. Foci formation was quantified using ImageJ (n=3). **(d)** Representative images are shown. As a control for BrdU incorporation, cells were treated with 2 M HCl to denature DNA prior to addition of the BrdU antibody. Scale bars; 10 µm.



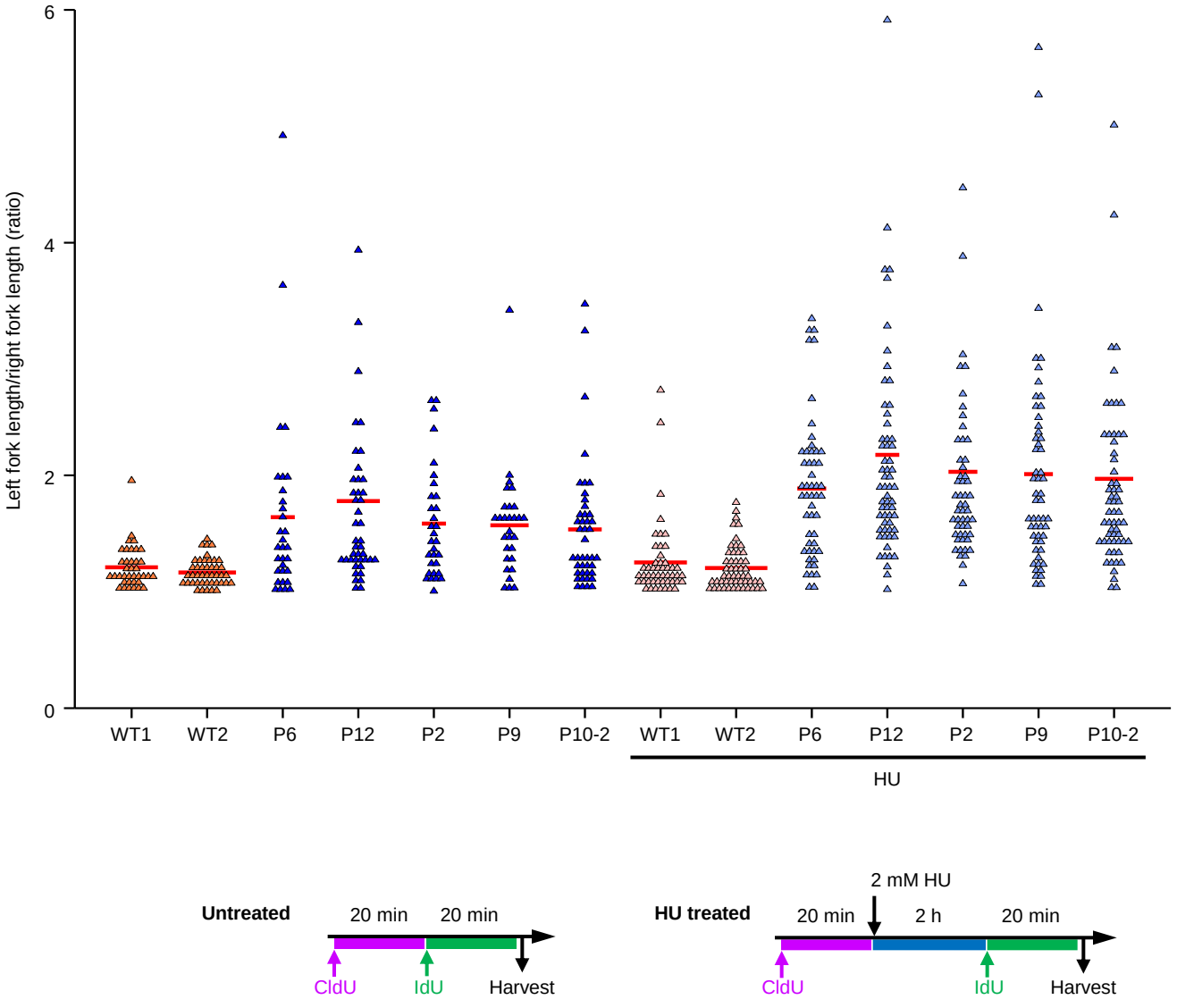

**Supplementary Figure 16: Replication stress exacerbates genome instability in DONSON deficient cells.**

Exposure of cells lacking DONSON to HU or MMC exacerbates micronuclei formation and chromosome breakage. **(a)** Micronuclei formation in HeLa cells transfected with either control or DONSON siRNA following exposure to either 4mM HU for 6 h, followed by 24 h recovery, or 50ng/ml MMC for 24 h was assessed using fluorescence microscopy (n=3). **(b)** Quantification of the chromatid gaps/breaks in metaphases from control or DONSON siRNA transfected HeLa cells following exposure to HU or MMC as in (a). At least 50 metaphases per experiment were counted (n=3). **(c)** Quantification of chromosome exchanges and radial chromosomes in metaphases from control or DONSON siRNA transfected HeLa cells following exposure to HU or MMC as in (a). At least 50 metaphases per experiment were counted (n=3). **(d)** Quantification of metaphases from control or DONSON siRNA transfected HeLa cells, containing highly fragmented or pulverised metaphases following exposure to HU or MMC as in (a) (n=3).



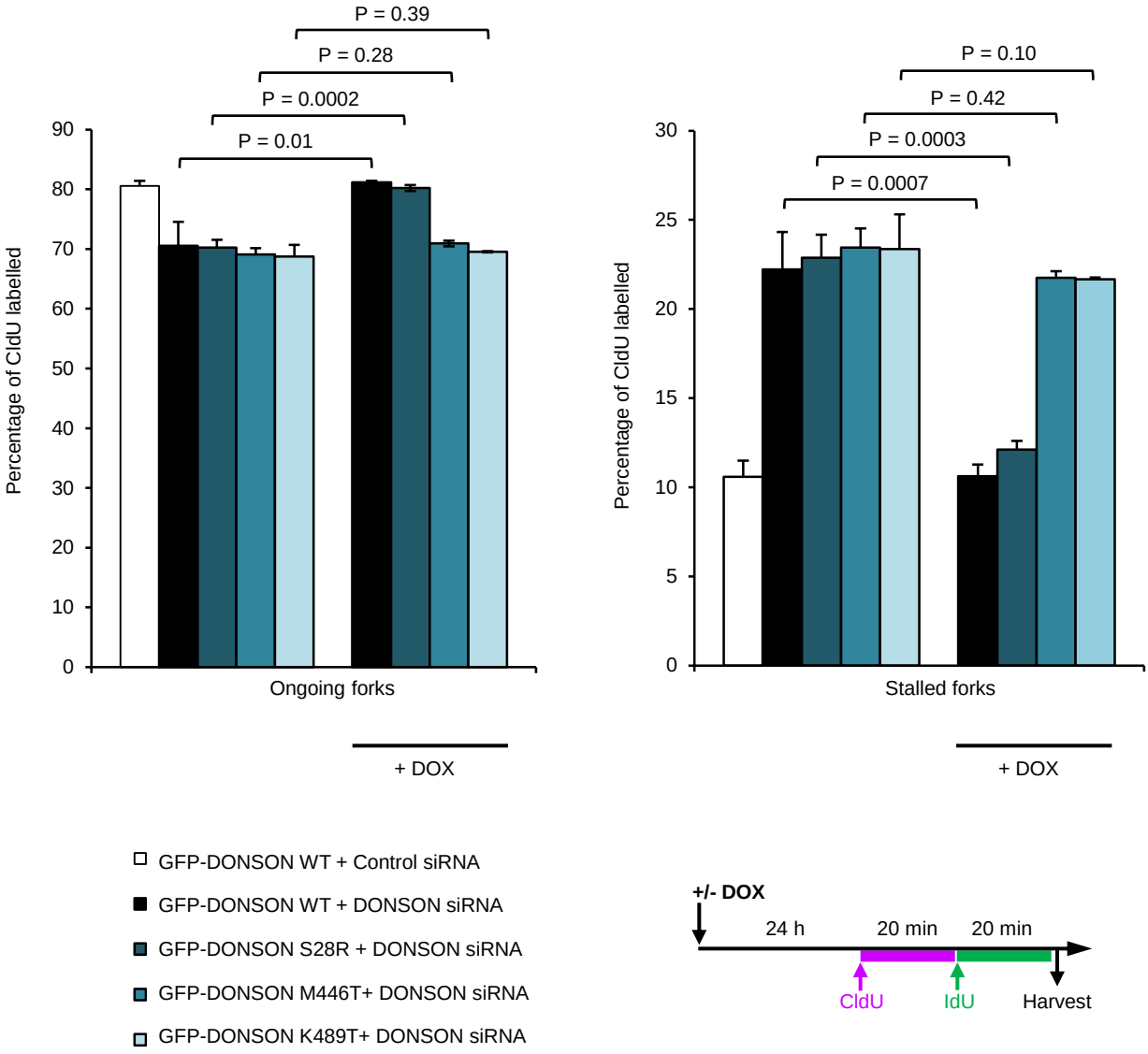

**Supplementary Figure 17: Cells from patients with hypomorphic *DONSON* mutations exhibit genome instability and replication fork abnormalities that are exacerbated by replication stress.**

**(a)** Mutation of DONSON results in replication fork instability that is exacerbated by replication stress. Patient derived fibroblasts were pulsed for 20 min with CldU, exposed to 2 mM HU for 2 h, and then pulsed with IdU for 20 min. DNA fibres were quantified, and the percentage of ongoing forks, new origin firing and stalled forks are displayed (n=3). Bottom right: Schematic of DNA fibre analysis. **(b)** DONSON patient derived fibroblasts exhibit elevated levels of spontaneous 53BP1/γH2AX foci. Patient derived fibroblasts were immunostained with antibodies to 53BP1 and γH2AX, and the percentage of cells with 53BP1/γH2AX foci were quantified (n=3). **(c)** Cells from DONSON patients exhibit spontaneous chromosomal instability. Chromosomal aberrations in patient derived fibroblasts were quantified. (n=3).



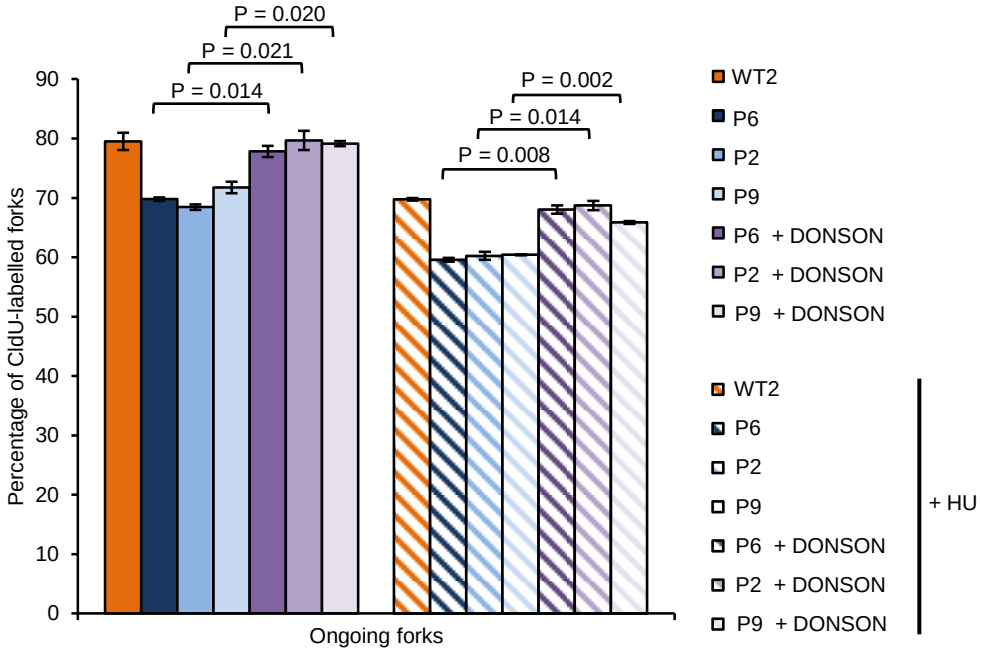

**Supplementary Figure 18: Cells from patients with hypomorphic *DONSON* mutations exhibit increased replication fork instability**

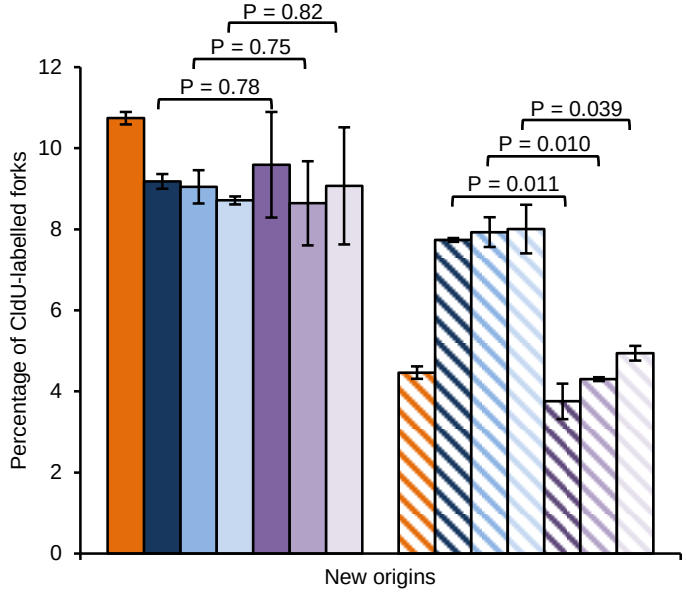
Quantification of DNA replication fork asymmetry in *DONSON* mutant patient-derived fibroblasts. The plot indicates the ratio of left/right fork track lengths of bidirectional replication forks. The red lines denote the median ratio (n=3). Bottom: Schematic of DNA fibre analysis.

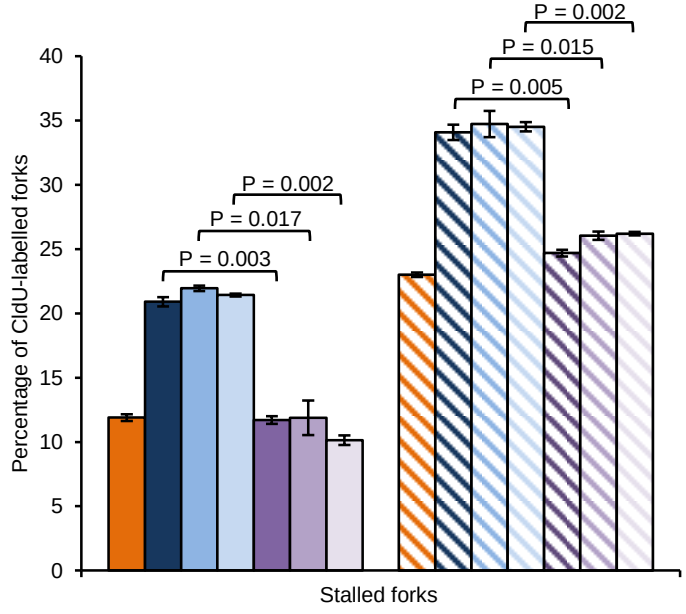


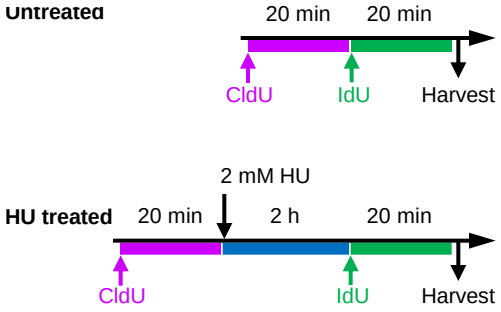

**Supplementary Figure 19: The patient-derived M446T and K489T mutations fail to correct the replication fork instability associated with DONSON depletion.**

HeLa Flp-In/T-Rex cells from (**Fig 2d**) expressing GFP-tagged WT or mutant DONSON under the control of a doxycycline-inducible promoter were transfected with DONSON siRNA. 48 h post-transfection, 1 µg/ml doxycycline was added for 24 h. The cells were pulsed with CldU for 20 min, followed by IdU for 20 min. DNA fibres were quantified, and the percentage of ongoing forks and stalled forks are displayed (n=2). '+DOX' indicates the addition of doxycycline 24 h prior to incubation with the thymidine analogues. Bottom: Schematic of DNA fibre analysis.



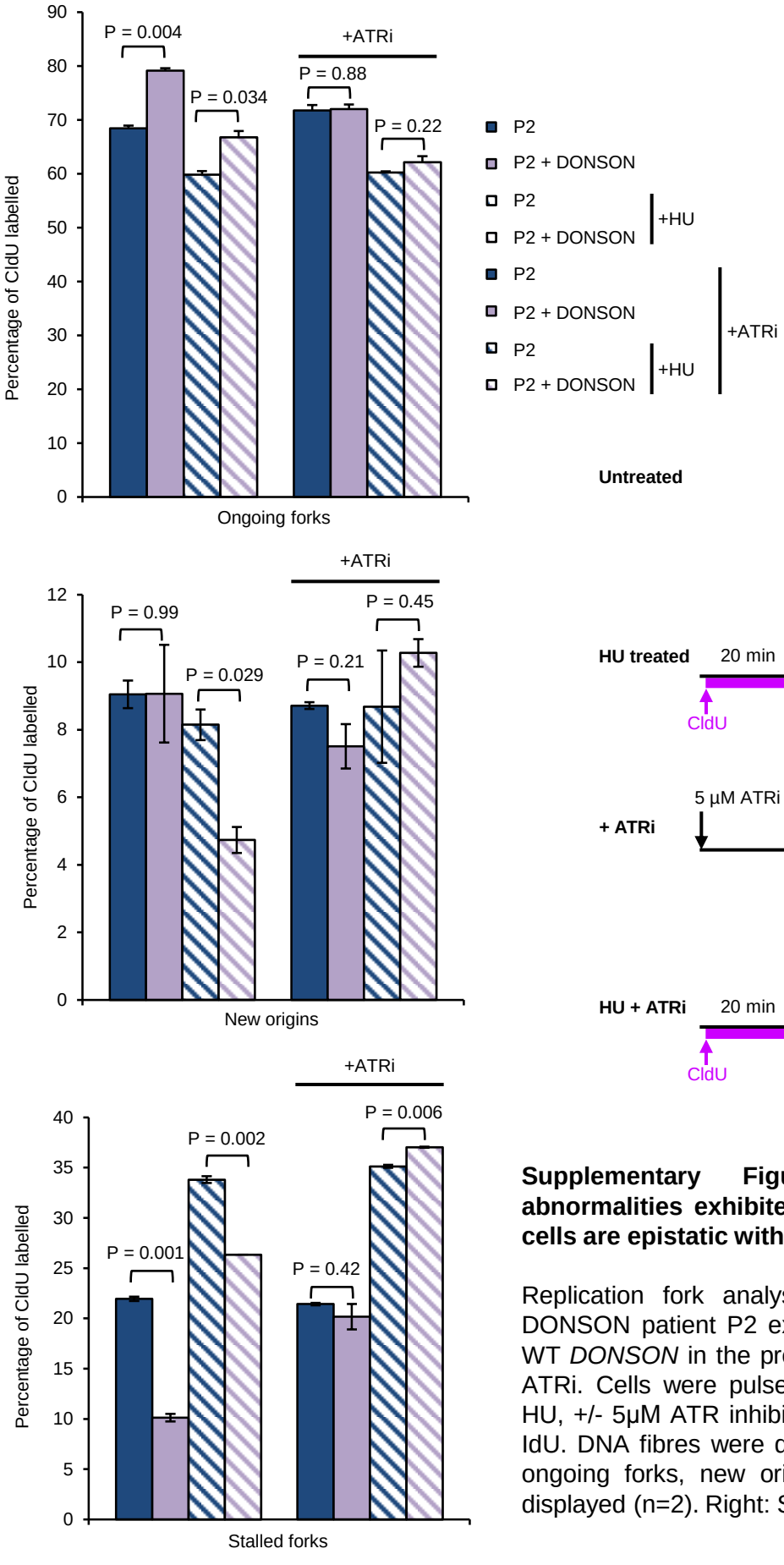



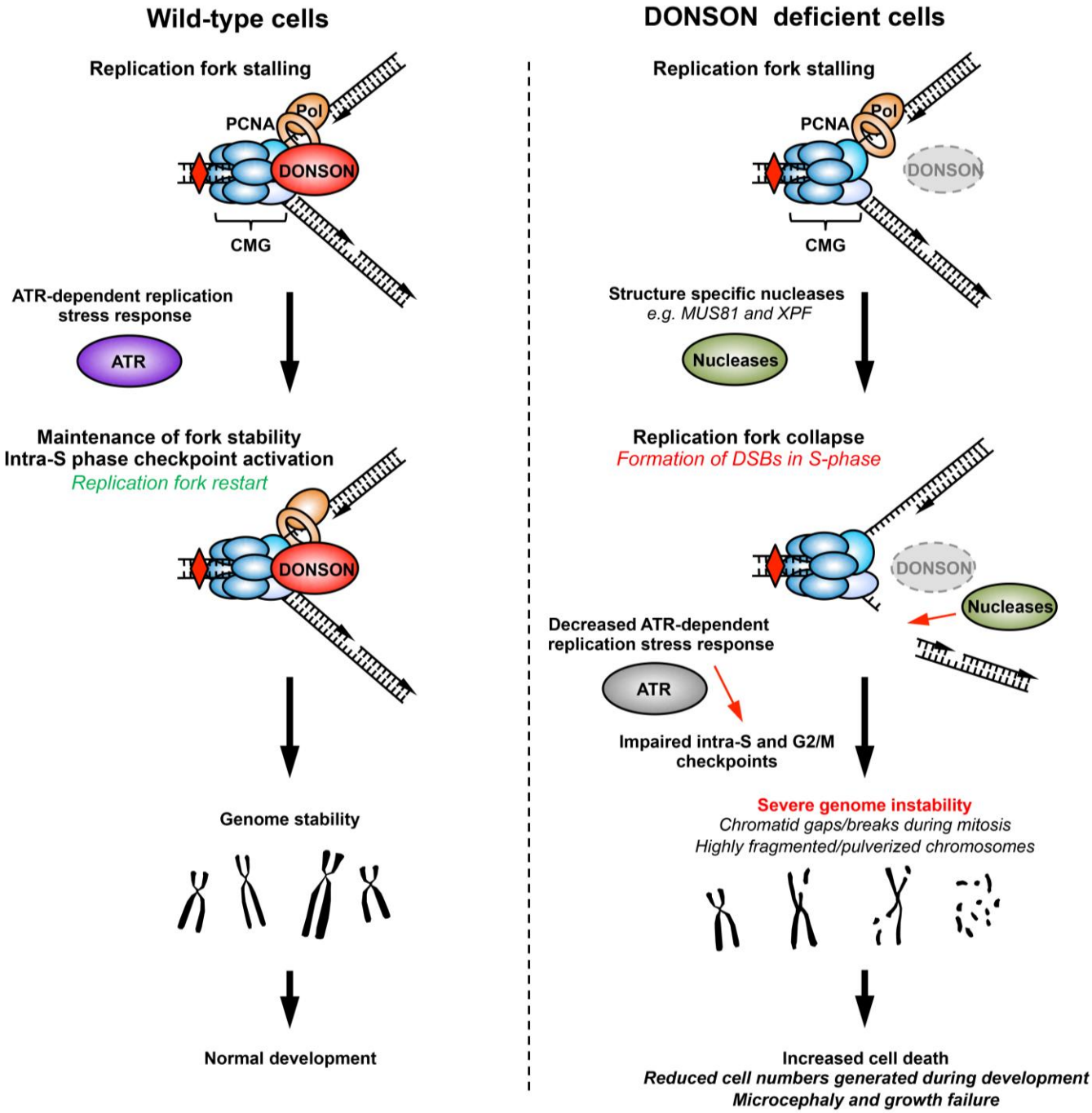

**Supplementary Figure 20: Re-expression of WT *DONSON* in patient derived fibroblasts restores replication fork stability.**

DNA fibre analyses in hTERT-immortalised DONSON patient fibroblasts complemented with a retrovirus either containing human WT *DONSON*, or an empty vector from (**Fig 7**). Complemented fibroblasts were pulsed with CldU, exposed to 2 mM HU for 2 h where indicated, and then pulsed with IdU. The percentage of ongoing replication forks, new origin firing and stalled forks were quantified (n=3 experiments). Quantification of new origin firing and stalled forks presented in (**Fig 7d**) have been included here for comparison. Right: Schematic of DNA fibre analysis.




**Supplementary Figure 21: The replication abnormalities exhibited by *DONSON* mutant patient cells are epistatic with ATR inhibition.**

Replication fork analysis of fibroblasts derived from DONSON patient P2 expressing either empty vector or WT *DONSON* in the presence or absence of HU and/or ATRi. Cells were pulsed with CldU, exposed to 2 mM HU, +/- 5µM ATR inhibitor for 2 h, and then pulsed with IdU. DNA fibres were quantified, and the percentage of ongoing forks, new origin firing and stalled forks are displayed (n=2). Right: Schematic of DNA fibre analysis.




**Supplementary Figure 22: Model depicting how DONSON mutations may lead to genome instability and microcephaly**

DONSON is a component of the replisome, a very large macromolecular complex of over two hundred proteins, including key replication machinery, the CMG complex (CDC45, MCM2-7 and GINS), PCNA, and the replicative polymerases (Pol). **(a)** Upon replication stress (red diamond) and replication fork stalling, the presence of functional DONSON ensures maintenance of fork stability. The ATR-dependent replication stress response also contributes to ensuring replication fork stability, and promotes activation of the intra-S and G2/M phase checkpoints, preventing transmission of DNA damage through to mitosis, and ensuring genome stability. **(b)** In the absence of functional DONSON, stalled replication forks are cleaved by structure-specific nucleases, such as MUS81 and XPF, leading to the generation of DSBs in S-phase. Impaired S-phase and G2/M checkpoint function in the absence of DONSON facilitates the transit of DNA damage into subsequent phases of the cell cycle, resulting in elevated mitotic chromosome breaks that will increase cell death. In patients with mutations in DONSON, such increased cell death would lead to reduced numbers of cells being generated during development, resulting in the microcephaly and growth failure observed.

**Supplementary Note:**

**Fluorescence Cross-Correlation Spectroscopy (FCCS)**

Fluorescence imaging and Fluorescence Correlation Spectroscopy (FCS) measurements were performed on a uniquely modified confocal laser scanning microscopy system, the ConfoCor3 (Carl Zeiss, Jena, Germany), consisting of the Zeiss LSM780 inverted setup and comprising Diode 405 nm, Ar multiline 458, 488 and 514 nm, DPSS 561 nm and HeNe 633 nm lasers. It enables detection using silicon Avalanche Photo Detectors (APDs) (SPCM-AQR-1X; PerkinElmer, USA) for imaging and FCS. Images were recorded at a 1024X1024 pixel resolution. The C-Apochromat 40x/1.2 W UV-VIS-IR objective was used throughout. Fluorescence intensity fluctuations were recorded in arrays of 10 consecutive measurements, each measurement lasting 10 s. Averaged curves were analyzed using the ZEN software for online data analysis or exported and fitted offline. In either case, the nonlinear least square fitting of the autocorrelation curve was performed using the Levenberg–Marquardt algorithm. Quality of the fitting was evaluated by visual inspection and by residuals analysis. Control FCS measurements to assess the detection volume were routinely performed prior to data acquisition, using dilute solutions ($10\,nM$) of Alexa488 and CF$^{TM}$568 dyes. Highly purified double-stranded DNA, in which each strand carries one fluorophore of Alexa488 and Atto565, was used as a cross-correlation standard. 85% cross-correlation was measured using the FCCS standard, whereas 13% cross-correlation (due to cross-talk) was measured with the two dyes in solution (data not shown). The variation between independent measurements reflects variations between cells, rather than imprecision of FCS measurements.

**FCS/FCCS measurements of GFP-DONSON/RFP-PCNA and GFP/RFP-PCNA expressing cells**

HeLa cells stably expressing GFP-DONSON and mCherry-PCNA (construct kindly provided by C. Lukas, Copenhagen; referred to as RFP-PCNA) were grown under standard conditions overnight on chambered coverslips (µ-slide, 8 well, Ibidi) and the growth medium was replaced by L-15 medium (Leibovitz) (Sigma) immediately prior to FCS/FCCS measurements. Cells expressing low levels of RFP-PCNA/GFP-DONSON or RFP-PCNA/GFP were used for FCS/FCCS measurements. Measurements were made in weakly expressing cells only. As the concentration of RFP-PCNA in replication foci was observed to be several-fold higher than that of DONSON-GFP and GFP, only cells with similar GFP/RFP ratios were used for cross-correlation analysis and comparable GFP-DONSON and GFP only concentrations used for comparison.

**Background on Fluorescence Microscopy Imaging and FCS**

As described above, an individually modified instrument (Zeiss, LSM780, ConfoCor 3) with fully integrated FCS/CLSM optical pathways was used for imaging. The detection efficiency of CLSM imaging was significantly improved by the introduction of APDs. As compared to PMTs, which are normally used as detectors in conventional CLSM, the APDs are characterized by higher quantum yield and collection efficiency – about 70 % in APDs as compared to 15 – 25 % in PMTs, higher gain, negligible dark current and better efficiency

in the red part of the spectrum. Enhanced fluorescence detection efficiency enabled image collection using fast scanning ($1-5\ \mu s/pixel$). This enhances further the signal-to-noise-ratio by avoiding fluorescence loss due to triplet state formation, enabling fluorescence imaging with single-molecule sensitivity. In addition, low laser intensities ($150–750\ \mu W$) could be applied for imaging, significantly reducing the photo-toxicity.

FCS/FCCS measurements are performed by recording fluorescence intensity fluctuations in a very small observation volume element (OVE) of a prolate ellipsoid shape (about $500\ nm$ wide and about $1-1.5\ \mu m$ long) that is placed in GFP-DONSON/RFP-PCNA or control GFP/RFP-PCNA HeLa cells by focusing the laser light through the microscope objective and by collecting the fluorescence light through the same objective using a pinhole in front of the detector to block out-of-focus light. We established that observed fluorescence intensity fluctuations are caused by fluorescently labeled GFP-DONSON or RFP-PCNA molecules passing through the OVE and that the characteristic decay times of the autocorrelation functions do not depend on the concentration of GFP-DONSON or RFP-PCNA (**Supplementary Fig. 13**). Therefore, we analyzed the diffusion of the two proteins in the nucleus, using temporal autocorrelation analysis.

In temporal autocorrelation analysis we first derive the autocorrelation function G(τ):

$$G(\tau) = 1 + \frac{\langle \delta I(t) \cdot \delta I(t+\tau) \rangle}{\langle I(t) \rangle^2},$$

where $\delta I(t) = I(t) - \langle I(t) \rangle$ is the deviation from the mean intensity at time $t$ and $\delta I(t + \tau) = I(t + \tau) - \langle I(t) \rangle$ is the deviation from the mean intensity at time $t + \tau$. For further analysis, an autocorrelation curve (ACC) is derived by plotting $G(\tau)$ as a function of the lag time, i.e. the autocorrelation time $\tau$.

To derive information about molecular numbers and their corresponding diffusion time, the experimentally obtained ACCs are compared to autocorrelation functions derived for different model systems. A model describing free three-dimensional diffusion of two components was used for fitting Cross-Correlation Curves (CCCs) in this study, whereas a two-component model with triplet correction was used for fitting ACCs:

$$G(\tau) = 1 + \frac{1}{N}\left(\frac{1-y}{\left(1+\frac{\tau}{\tau_{D_1}}\right)\cdot\sqrt{1+\frac{w_{xy}^2\tau}{w_z^2\tau_{D_1}}}} + \frac{y}{\left(1+\frac{\tau}{\tau_{D_2}}\right)\cdot\sqrt{1+\frac{w_{xy}^2\tau}{w_z^2\tau_{D_2}}}}\right)\cdot\left(1+\frac{T}{1-T}\cdot e^{-\frac{\tau}{\tau_T}}\right)$$

In the above equation, *N* is the average number of molecules in the OVE; *y* is the fraction of the slowly moving GFP-DONSON and RFP-PCNA molecules; $\tau_{D_1}$ is the diffusion time of the free GFP-DONSON and RFP-PCNA molecules; $\tau_{D_2}$ is the diffusion time of GFP-DONSON and RFP-PCNA molecules undergoing interactions; $w_{xy}$ and $w_z$ are radial and axial parameters, respectively, related to spatial properties of the OVE; T is the average equilibrium fraction of molecules in the triplet state; and $\tau_T$ the triplet correlation time related to rate constants for intersystem crossing and the triplet decay. Spatial properties of the detection volume, represented by the square of the ratio of the axial and radial parameters ($\left(\frac{w_z}{w_{xy}}\right)^2$), are determined

in calibration measurements performed using a $10\,nM$ solution of Alexa488 and CF$^{TM}$568 for which the diffusion coefficient (D) is known. The diffusion time, $\tau_D$, measured by FCS, is related to the translation diffusion coefficient *D* by:

$$\tau_D = \frac{w_{xy}^2}{4D}.$$